DOCUMENT RESUME

ED 142 932 CS 003 396

AUTHOR	Hakes, David T.
TITLE	Psychological Processes in Sentence Comprehension. Final Report.
INSTITUTION	Texas Univ., Austin. Dept. of Psychology.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C. Office of Research.
BUREAU NO	2-0216
PUB DATE	Dec 74
GRANT	OE-G-0-72-3941
NOTE	324p.
EDRS PRICE	MF-$0.83 HC-$16.73 Plus Postage.
DESCRIPTORS	*Ambiguity; *Comprehension; Comprehension Development; Cues; *Idioms; *Language Development; Language Research; Linguistics; *Psycholinguistics; *Sentences; Sentence Structure; Theories

ABSTRACT

This report summarizes two years of research by a team at the University of Austin, aimed at developing and testing a model of sentence comprehension. The general model that has guided the work maintains that, on the basis of cues available in a sentence's surface form, the hearer projects hypotheses about the grammatical relations and semantic content that underly the sentence. Separate sections of the report discuss research related to the nature of the surface cues and to the ways such cues are used during comprehension; the processing of ambiguous sentences; and the comprehension of idioms, which pose a number of linguistically and psycholinguistically interesting problems. Specific experimental data are reported in a series of extensive appendixes. (AA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED142932

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL REPORT

Project No. 2-0216

Grant No. OE-G-0-72-3941

PSYCHOLOGICAL PROCESSES IN SENTENCE COMPREHENSION

David T. Hakes
Department of Psychology
The University of Texas at Austin
Austin, Texas 78712

December, 1974

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

National Institute of Education
Office of Research

CS 003 396

I. Introduction

The research reported here has been guided by a model of comprehension processing which maintains that on the basis of cues available in a sentence's surface form the hearer projects hypotheses about the grammatical relations and semantic content which underly the sentence. This heuristic strategy model of comprehension processing has been presented in some detail elsewhere, (see, e.g., Bever, 1970; Hakes, 1972a). Since the research presented here forces no substantial alterations in the model as it has been presented, the model itself will not be detailed again here.

It is of some interest to note in passing that recent attempts to construct computer-based comprehension devices (e.g., Winograd, 1972) have adopted much the same heuristic strategy approach to comprehension processing that has informed our research. It is not clear, however, that the development of "comprehending computers" is of any great value in advancing our knowledge of how the human hearer accomplishes the comprehension task. One reason for this is simply that as yet we know far too little about the nature of the heuristics the hearer uses or about the nature of the processor that uses them. It becomes painfully apparent in examining the empirical data available on comprehension processing, including those presented here, that we have only just begun the enormously complicated task of discovering the kinds of information available to the human comprehender and the kinds of uses he makes of them.

In a very real sense, there does not as yet exist anything that deserves to be called a model of the comprehension process. The "model" that does exist is, in reality, little more than a sketch of the general form we believe an adequate comprehension model will ultimately take.

One consequence of this state of affairs is that there exist relatively few well-motivated constraints on the detailed functioning of such a model. But computers are notoriously intolerant of vagueness, requiring that we tell them extremely explicitly how to do what we ask them to do. Thus for all those aspects of comprehension processing that are unconstrained either by the nature of the language itself or by what little we know of the workings of the human comprehender the computer model builder must guess. And there is no assurance that the model that results from making large numbers of such guesses, even if it will "comprehend" some sentences, does so in ways anything like those employed by the human.

Unfortunately, the situation is even worse than this in several ways. For one, the ad hoc assumptions the model builder must incorporate in order to get the program to run at all are so deeply embedded in that program that it is difficult, if not impossible, to isolate any one of them and examine its effects on the total functioning of the comprehension program. Put differently, the assumptions about what the heuristics are and how they are used are nearly impossible to discover from examining the program itself. And the result of this is that there is great difficulty in designing experiments to determine whether the heuristics postulated for the computer are reasonable approximations of those used by the human.

For these reasons, we have found it relatively unprofitable to study the existing computer-based models of comprehension processing. While we have found it encouraging that the people concerned with such models have adopted the same general theoretical approach to the problem as ours, our inspiration has had to come from other sources.

It is perhaps also worth noting that there has in the recent past been relatively little research of direct relevance to problems of the nature of comprehension processing per se. In reflecting on this dearth of relevant research from other laboratories, two kinds of reasons become apparent. The first is simply that the vagueness of current models of the process leaves it very unclear what the possible heuristics and processes might be, providing little direction for such research. We have found in our own research that finding reasonable hypotheses about the kinds of cues and strategies that the human comprehender might use and about the nature of the processes themselves generally requires lengthy sifting of sentences and non-sentences, looking for features that intuitively seem to cause comprehension difficulty or, alternatively, appear not to cause such difficulty when for one reason or another we would have expected it.

There appears to be one area which counts as an exception to the claim about a dearth of research pertinent to comprehension processing--the area of developing comprehension in children. Much of Bever's (1970) original discussion of comprehension strategies was couched in terms of the development of strategies, and many of the data on which he relied were developmental data. Since that time there has been considerable discussion of strategies for comprehension in young children (see, e.g., Slobin, 1973) and also a considerable amount of research that can be interpreted in terms of the development of comprehension strategies (see, e.g., Sinclair & Bronckart, 1972; Sheldon, 1974).

There appears, unfortunately, to be no simple relation between the comprehension performance of the children and that of adults. There is, of course, no a priori reason to believe that there should be a simple relation between what the child does, operating with an incomplete and changing set of comprehension strategies, and what the adult does, presumably operating with a nearly complete set of strategies. We will consider later in some detail a case in which there appear to be marked discrepancies in the patterns of comprehension for a set of related constructions between the children and adults (see below, Section II).

The second reason for the dearth of recent research directly concerning comprehension processing appears to be that psycholinguists interested in comprehension have been devoting far more attention to the products of comprehension processing than to the processing itself. In this regard, two recent developments have captivated the attention of psycholinguists. One is the recent series of demonstrations by Bransford, Franks, and their associates (e.g., Bransford & Franks, 1971) that when presented a series of related sentences hearers tend to integrate information across sentences, often inferring information that was nowhere stated explicitly. In addition, Sachs (1967, 1968) and others have demonstrated that within a few seconds after hearing a sentence little information is available about the detailed structural characteristics or lexical items of the sentence. These two kinds of demonstrations, taken together, suggest that the product with which the comprehension process ends is a highly abstract one, only remotely related to specific characteristics of the input. That this should have come as a surprise to anyone is itself surprising. But demonstrating that the representation is abstract and "semantic" has combined with a second trend to refocus the field's attention.

This second trend, promulgated by linguists and philosophers, has involved changing and broadening the scope of linguistic inquiry from a dominant concern with syntax to a concern with the semantic and pragmatic aspects of language. Even a casual review of recent literature reveals a plethora of material on the literal and conveyed meanings of sentences, their entailments and the inferences they invite, their semantic and/or pragmatic presuppositions, and the implicatures that arise from their use in one or another conversational setting. And while it is not yet clear that the new "semantic" focus of linguistics has done much more than name a new set of problems, what is clear is that such problems have become immensely attractive to psycholinguists. Not surprisingly, we have entered upon another era of "psychological reality" experiments. But while the psychological reality experiments of the '60's were focused on the reality of various syntactic structures and rules, the corresponding experiments of the '70's are focused on the reality of semantic and pragmatic constructs (see, e.g., Clark, 1973; Perfetti, 1972).

On first consideration it would seem reasonable to suppose that if conceptions of the ultimate product of comprehension processing change then conceptions of the processing itself must correspondingly change. For how can we say how to get from A to B without knowing where B is? If this were the case, then hope of making any progress on an account of comprehension processing would be dim indeed; for as yet there is little more to be said about the representation that results from comprehension processing other than that it is only abstractly related to the input. The problems involved in finding an adequate characterization for that representation are truly monumental and will, no doubt, absorb the efforts of linguists, psycholinguists and philosophers of language for a long time.

But the prospects for making progress toward an account of comprehension processing are not as bleak as this would suggest. It may be true that we cannot say how the hearer constructs a full, abstract, semantic representation of what he has heard until we have some reasonable idea of what such a representation should look like. We can, however, be reasonably certain that certain kinds of information must be recovered from the input as necessary prerequisites for constructing such a representation. Consequently, while we may be very uncertain about all the kinds of information that are recovered, we may at least be able to say something about the recovery of those that we do know something about.

Consider, for example, the overworked case of active and passive sentences. It is now generally accepted that the import of such sentences is slightly different, a difference spoken of by some as one of "theme", by others as a difference of "focus", and by still others in still different ways. It is unclear whether this is a difference in meaning per se or whether active and passive sentences are fully synonymous but differ in other ways (as, e.g., Katz, 1972, would have it). And it is also unclear how this difference is to be represented in the representations hearers construct upon hearing such sentences. But what is clear is that there is some sort of difference and, further, that that difference must be recoverable from differences in the superficial forms of the two kinds of sentences. The difference clearly has something to do with the roles of the two NPs in the sentences; and whether we characterize those roles syntactically (i.e., "subject" and "object"), semantically (e.g., "agent" and "patient") or in some as yet undiscovered way, that difference must be keyed by either information recoverable about the characterization of the nouns occurring in the sentences or from such superficial cues as whether or not there is something like was + (V)-ed + by wrapped, so to speak, around the verb.

Thus, to the extent that we know that some sort of difference in structure or meaning must be recoverable from differences between types of sentences, we can make some progress on problems of what superficial differences cue those differences in underlying representation and how this is accomplished even if we do not know how to characterize the underlying representation of those differences. In practice we have found it far more useful to consider the linguist's intuitions about where differences lie than to consider their theoretical accounts of how those differences are to be represented.

But while some progress can be made toward developing a model of the comprehension process without waiting for fuller development of an adequate way of representing the products of that process, such progress is painfully slow and difficult. The remainder of this report is devoted to discussing the research we have performed in the past two years in the interests of developing and testing a model of the comprehension process.

For purposes of convenience, the research has been divided into three sections, even though the theoretical considerations involved overlap greatly among these. Section II deals with research directly related to the nature of surface cues to underlying structure and meaning and to the testing of some hypotheses about particular ways such cues are used during comprehension. Section III concerns the processing of ambiguous sentences and with the special issues raised by ambiguity. Section IV concerns the comprehension of idioms. Idioms pose a number of linguistically and psycholinguistically interesting problems, but ones that have only rarely been considered in the context of comprehension processing. While we can hardly claim to have resolved any of the problems, our results do place some interesting constraints on what the solutions could be.

II. Comprehension Strategies and Surface Cues.

A. Research on Verb Structure

For the past few years much of our attention has been absorbed in examining the effects of sentential complexities associated with verbs. Our interest was aroused by a paper by Fodor, Garrett and Bever (1968) in which it was argued that the main verb of a clause plays an important and unique role in comprehension processing. They hypothesized that a hearer, on encountering a verb, retrieves from his mental lexicon information about the predicate structures that could possibly be associated with that verb. This information forms the basis for projecting a set of structural hypotheses about the material which might follow that verb, hypotheses which are then tested against further incoming information to discover the predicate structure actually realized in the sentence being processed.

Fodor et al are not alone in attributing a prominent place to verbs. Generative semantics (e.g., Lakoff, Ross, Postal) have tended to treat verbs as representing the basic predicates of the propositions underlying sentences, the noun phrases, etc. associated with them being represented as arguments of those predicates. This sort of representation appears, for example, in the "case" grammar of Fillmore (1968) where it is the verb that determines the number and variety of arguments the proposition may have. The notion that it is a sentence's verb that is the main determinant of the sentence's structure and meaning has also surfaced in the psycholinguistc literature (e.g., Healy & Miller, 1970, 1971).

Given the important role accorded to verbs. it would hardly be surprising if they played a central part on comprehension processing. A further indication of the possible nature of this role comes from an observation made by Robert Kirk (personal communication). Kirk has noted that the possible structures of English and of other languages are constrained so as to provide information early in any clause as to whether that clause is main or subordinate and that this is information essentially about that clause's main verb. The constraint is, essentially, that if a subordinate clause precedes a sentence's main clause, the subordinate clause must contain cues occurring no later than that clause's main verb (or, in the case of verb-final languages, no later than the auxilary verb) indicating its subordinate status.

But even if verbs play a central role in comprehension processing there remains a great variety of possibilities for the nature of that role. Fodor et al proposed one such role--that information about the verb's possible predicate structures is used in an active, predictive way. From this they derived the hypothesis that sentences containing verbs that could enter into many different predicate structures should be more difficult to comprehend than sentences whose verbs could enter into only a few structures.

Fordor et al reported two experiments testing this "verb complexity" hypothesis. In the first, they used doubly self-embedded sentences which differed only in whether one of their clauses contained a verb which required a simple NP direct object (i.e., a transitive verb) or a verb which could take either that predicate structure or a sentential complement as direct object. In both cases, the sentences actually contained simple NP objects, e.g., (1) and (2). Subjects were required to paraphrase the

(1) The letter the secretary the manager employed mailed was late. (Transitive)

(2) The letter the secretary the manager employed expected was late. (Complementizing)

sentences, and the results indicated that the sentences were paraphrased more accurately when they contained simple, transitive verbs than when they contained complex, complementizing verbs. In the second experiment, similar pairs of sentences were used in an anagram solving task. Some of the results, though not all, were consistent with the verb complexity hypothesis.

Hakes (1971) noted that the verb complexity hypothesis, by implying an active, on-line use of information about possible predicate structures, implies that differences in verb complexity should arise soon after the verb occurs and, hence, should be detectable in an on-line processing task. Two experiments were reported there testing this. In both, paraphrasing and phoneme monitoring data were collected for sentences like (3) and (4).

(3) When he had heard all sides of the dispute, the manager blamed the foreman who had supervised the job. (Transitive)

(4) When he had heard all sides of the dispute, the manager believed the foreman who had supervised the job. (Complementizing)

In the first experiment the phoneme monitor target was either the verb itself or the head noun of the verb's object NP; in the second experiment it was always the head noun of the object NP. The paraphrasing data from the first experiment replicated the effect reported by Fodor et al, i.e., paraphrasing was less accurate when the sentences contained complementizing verbs than when they contained transitive verbs. However, the paraphrasing effect was not obtained for the second experiment. More important, no effect of verb complexity was apparent in the phoneme monitoring data for either experiment.

Thus, the data of our first experiments were mixed. The lack of a complexity effect in the on-line phoneme monitor task suggested that Fodor et al's account of their paraphrasing data was incorrect, that information about verb predicate structure was not being used in the manner they suggested. The paraphrasing data, on the other hand, suggested that verb complexity did have some effect on comprehension processing. Since then, we have performed three additional experiments designed to test a number of alternative hypotheses about when and how information about verb structure was retrieved and used. These experiments have been described in considerable detail in a recent unpublished paper which is included as Appendix I of this report.

The results of those experiments are somewhat complicated, but in general their results are consistent with those of the earlier experiments: increased verb complexity generally results in decreased paraphrase accuracy, but appears to have no concsistent effect on phoneme monitoring latency, regardless of where the monitoring target is located. Why this should be the case is not at all clear. There is, however, some additional evidence that suggests that the differences between the effects obtained with on-line comprehension tasks and those obtained on post-sentential tasks are real differences and are not dependent on the particular on-line task used in our research. Chodorow and Garrett (personal communication) have been unable to find any consistent verb complexity effects using an on-line task other than phoneme monitoring. Their research involved an on-line probe task in which, while the subject is listening to a sentence, he either hears or sees a single word. His task is to indicate, as quickly as possible, whether or not that word has occurred in the sentence he is hearing.

What is to be made of the on-line probe results is somewhat unclear, for it appears that this task does not always yield results comparable to those of the phoneme monitoring task, a point to which we will return later. However, the failure to find verb complexity effects with either the phoneme monitoring or on-line probe tasks could mean that the effects found in paraphrasing do not arise during comprehension processing. It seems possible that there is no greater difficulty recovering the underlying structure and meaning of a sentence when that sentence contains a verb that could have occurred in a variety of other sentence structures than when it contains a verb that could occur only in the structure in which it did occur. If this is what the results obtained with on-line comprehension tasks are telling us, then it seems most likely that the paraphrasing effects indicate something about the processing involved in constructing a paraphrase for a sentence after that sentence has been understood. How it might be that the other, unused readings of a verb might interfere with this process is an open question, for there exists no account of how paraphrases are constructed. But that there might be such interference seems at least intuitively plausible.

Although we have spent considerable time considering the problem from this perspective, and from a number of others as well, we have been unable to develop any account that accords well with the available data on verb complexity and also leads to interesting empirical predictions. There are, for example, too few threads available on which to build an account of the processes involved in constructing and producing a sentence as a paraphrase of another sentence. And while we have had some thoughts along those lines, they have been far too fragmentary to be empirically interesting. So, in the absence of any clear indications of how to pursue the verb complexity problem further, we have declared ourselves (temporarily) mystified and turned our attention elsewhere.

B. Research on Relative Clause Structure and Reduction

One focus of that attention has been on sentences containing relative clauses. Our interest in such sentences was stimulated several years ago by a series of experiments by Fodor and Garrett (1967). They advanced the hypothesis that the relative pronouns that introduce relative clauses are used by the hearer as cues to the fact that the sentences have such clauses embedded in them. From this, they reasoned that if such relative pronouns were deleted the sentences would be more difficult to comprehend. To test this, they presented their subjects with doubly self-embedded sentences, e.g., (5), which differed only in whether the relative pronouns

(5) The lion (which) the tiger (that) the gorilla chased killed was ferocious.

were present or absent. The subjects were required to paraphrase each sentence, and the results indicated that the paraphrases were less accurate when the pronouns were deleted than when they were present.

These results suggested that the relative pronouns are important cues to the sentences' underlying structure and imply that such cues are used during comprehension processing, for the structure and meaning to be recovered are the same, regardless of whether the pronouns are present or absent. But beyond this Fodor and Garrett's data shed little light or when or how such optional cues are used.

Later experiments by Hakes and Cairns (1970) and Hakes and Foss (1970) indicated that the effect of relative pronoun deletion could be detected during comprehension processing. They showed that phoneme monitoring latency was longer for doubly self-embedded sentences from which the relative pronouns had been deleted than for the same sentences with the relative pronouns present.

At that point, several things remained unclear. First, the data did not resolve the question of when during processing the difference arose. Both the Hakes and Cairns and the Hakes and Foss experiments measured monitoring latency late in the sentences, i.e., on the second of the sentences' three verbs. This left open the question of whether the difficulty difference arose immediately upon the occurrence of the relative pronoun (or upon its failure to occur) and was a sufficiently long-lasting effect that the difference could still be detected considerably later in the sentence or whether there was a considerable lag between the occurrence of a cue and the occurrence of the processing using that cue. What would happen, that is, if monitoring latency were measured much closer to the relative pronoun, e.g., within the relative clause itself?

A second unanswered question was whether the increased comprehension difficulty that resulted from deleting an optional cue occurred only when the sentence was difficult for other reasons. That is, is there an interaction between "general" processing load and the "local" load attributable to deletion of an optional cue such that the latter is negligible except when the former is great? Further, the available data left unclear whether the cue deletion effect was really that, i.e., an effect that accompanied the deletion of any optional cue, or whether it was an effect specific to relative pronouns.

This last is a difficult question to answer, for there appear to be few cases where there is an optional element that can be deleted without affecting the semantic interpretation of the resulting sentence. One such case is that of the optional complementizer that in sentences containing NP predicate complement constructions, e.g., (6). Hakes (1972)

(6) The blind student felt (that) the material in the art course would be too difficult for him to understand.

performed two experiments on the deletion of this cue, using both the phoneme monitor and paraphrase tasks, and finding that here, too, the deletion increased comprehension difficulty. Since the sentences were somewhat less dificult to understand than the double self-embeddings that have been used in studying relative pronoun deletion, these results suggest that the increase in comprehension difficulty resulting from cue deletion does not crucially depend on a high level of overall difficulty. In addition, the monitor target in both experiments was the head noun of the subject NP of the embedded complement, e.g., the /m/ in material for (6), or a prenominal adjective in that same NP. Since the distance between the optional cue and the target was small, it appears that however such optional cues are being used their effect lags little behind their occurrence.

While these results were suggestive of answers to some of the qeustions we have raised, we felt the questions merited further study. In the course of thinking about embedded relative clauses and about how the relative pronoun cue might be utilized, we encountered another question, one about why self-embedding of relative clauses created so much comprehension difficulty.

The doubly self-embedded sentence is the sort that psycholinguists and linguists frequently cite as a type of sentence so difficult to comprehend as to be nearly impossible. The "standard" account of this difficulty is one suggested by Miller (e.g., Miller & Chomsky, 1963; Miller & Isard, 1964) and others and is based on the fact that a self-embedded relative clause modifies the subject of the clause in which it is embedded. As a result, the embedded clause occurs between the subject and predicate of the embedding clause, interrupting that clause. Miller's account of the difficulty attributed it to this interruption of the on-line processing of the embedding clause.

In reviewing the literature on the comprehension of self-embedded sentences, we became aware that every experimental test of the interruption hypothesis that had been performed was confounded, including the one which we ourselves performed (Hakes & Foss, 1970). Essentially, the comparison has always been between doubly self-embedded sentences and sentences with two right-branching relative clauses, e.g., (7) vs. (8).

(7) The car that the man that the dog bit drove crashed.

(8) The dog bit the man that drove the car that crashed.

There does seem to be some difference between such sentences in what they are about, for the first is a sentence about the car and what happened to it (and why), while the second is about the dog, what he did, and what effect that had. More striking than this is, of course, the difference in the difficulty of the two, a difference so obvious that experiments are hardly necessary to establish it.

What attracted our attention, however, is an additional difference. The relative clauses in the self-embedded sentence are object relative, i.e., ones in which it is the direct object of the clause that is relativized. In the right-branching sentence, on the other hand, the relative clauses are subject relatives, one in which it is the clause's subject that is relativized. This difference in the internal structure of the relatives results in a difference in the order in which their constituents appear in the superficial form of the sentence, for an object relative involves fronting the direct object as well as relativizing it.

We wondered if this difference in the internal structure of the relative clauses might have something to do with the difference in difficulty. As a game, we constructed sentences containing two object relative clauses that modified the object rather than the subject of the enbedding clause, i.e., sentences where the embedding clause was not interrupted, e.g., (9). And our intuitions told us that these, too, were quite

(9) The car was driven by the man that the cat that the dog scared scratched.

difficult to understand, even though they involved only a single level of self-embedding (i.e., of one relative clause within the other). Thus we were led to suspect that the source of the difficulty might lie in the internal structure of the relative clauses rather than in whether or not they interrupted an embedding clause.

With this possibility in mind, we designed two experiments to study the effects of relative clause structure and location on comprehension difficulty. In addition, in response to the questions raised earlier about relative pronoun deletion, we included deletion as a variable in both experiments. The two are described in detail in Appendix II, and we will not repeat that information here. The outcomes of the two experiments were quite in accord with our intuitions. In the first experiment, which compared object relative clauses that were either self-embedded or right-branching, the results indicated that self-embedding per se was not a major factor contributing to comprehension difficulty. That is, comprehension difficulty was actually greater, as indicated by both the phoneme monitoring and the paraphrasing results, when the object relative clauses were right-branching than when they were self-embedded. Clearly, this is not the sort of result one would expect on the basis of the interruption hypothesis.

The results of the second experiment, which compared subject relative clauses that were either right-branching or self-embedded, also lent no support to the interruption hypothesis. Here there was no difference in difficulty between clauses in the two different locations. Although there are substantial reasons for discounting comparisons between the results of the first and second experiments, reasons detailed in Appendix II, the overall pattern that the two suggest is that subject relatives result in less comprehension difficulty than object relative regardless of whether they are self-embedded or right-branching. Right-branching object relatives result in greater comprehension difficulty than self-embedded object relatives.

Thus it would appear that the results are entirely consistent with the hypothesis that the major source of difficulty in comprehending sentences containing relative clauses is the internal structure of the relative. Apparently, changing the order in which the subject, verb and object of the relative clause occur from the order in which they occur in simple main clauses--a change that occurs in object relatives but not in subject relatives--contributes to comprehension difficulty. Thus our results suggest that the canonical-order comprehension strategy suggested by Bever (1970)--assume that constituents will occur in the usual order (NVN=subject/Verb/Object) unless otherwise indicated--is a dominant strategy guiding sentence comprehension.

But if adults use this canonical ordering strategy, an apparent paradox emerges. There is abundant evidence that sensitivity to constituent ordering is one of the first conprehension strategies to emerge in children (see Bever, 1970; de Villiers & de Villiers, 1973, for some of the relevant evidence). But if children are using such a canonical ordering strategy before they begin understanding sentences containing relative clauses, then we should expect children to show the same pattern of comprehension difficulty results for sentences containing self-embedded or right-branching subject or object relative clauses that we have found for adults. Yet all the available evidence suggests that this is not the case.

Sheldon (1974), for example, has performed a comprehension experiment with children between the ages of 3 and 5 years, using the same four combinations of relative clause structure and location that we used. The task was, of course, quite different, the children being given a set of toy animals with which they were to act out the events described in the main and relative clauses of each sentence. Sheldon's results indicate that right-branching object relative clauses are more easily comprehended by the children that self-embedded relative clauses. However, self-embedded subject relative clauses are understood better than right-branching relatives.

Sheldon suggests on the basis of both data on accuracy of comprehension and data on the nature of the errors the children made that the children seem to be assuming that the relativized item in the relative clause serves the same grammatical function in that clause that it serves in the main clause. That is, when a relative clause is self-embedded (i.e., modifies the subject of the main clause), the children apparently assume that the relative pronoun will be the subject of the relative clause.

Similarly, when the relative is right-branching (i.e., modifies the object of the main clause), they apparently assume that the relative pronoun is the direct object of the relative clause as well as the direct object of the main clause.

Brown (1971), using a picture-selection comprehension task with children between 3 and 5 years, has reported results that appear to show the same pattern of comprehension difficulties as Sheldon's results. (In fact, the form in which Brown has presented his results makes it a bit difficult to determine whether this is in fact the case.)

It may be that the difference in the pattern of results obtained with children from that we have obtained with adults is attributable to differences in the kinds of tasks used. Some support for this possibility comes from a recent report by Smith (1974) of preliminary results obtained from children using an elicited imitation task, for Smith's results are not consistent either with ours or with those of Sheldon and Brown. Smith's results seem more consistent with the interruption hypothesis than any of the others.

But if the difference between the child and adult results is not wholly a methodological artefact, then the difference suggests an interesting possibility about the child's acquisition of comprehension strategies(and/or a grammar). It is generally assumed, though seldom explicitly, that the child's set of comprehension strategies (or grammatical rules) is a subset of the adult's. The progress the child makes is viewed as the result of his adding strategies or rules so that his incomplete set becomes progressively more complete until he arrives at the adult's set. Thus, at the stage where the child understands active sentences but fairly consistently misunderstands passive sentences, it is assumed that the child has already mastered the canonical order strategy and consistently treats the first NP in a sentence as the subject of the verb that follows. It is not until later that he discovers that passives, signalled perhaps by the presence of by between and main verb and the following NP, constitute an exception to this strategy and require an additional strategy for their correct interpretation.

The relative clause results suggest, however, that the child's set of comprehension strategies may not be just a subset of the adult's. Rather, they suggest, as Sheldon has noted, that at least some children adopt the strategy of interpreting the relativized NP as serving the same grammatical function in both main and relative clauses, a strategy Sheldon calls the "parallel function hypothesis." Such a strategy cannot be a part of the adult's set of comprehension strategies; for although it is perhaps reasonable that the English language might incorporate some such constraint on relative clause structure and interpretation, it is clear that there is no such constraint. Thus, if children are in fact trying out such a strategy for interpreting relative clauses, they are trying out one that does not work and which they will have to abandon. The import of all this is just that the child's comprehension strategies may not be limited by the adult's, as is usually assumed, and instead may from time to time contain some plausible but incorrect ones. If so, then the child's progress toward understanding language the way an adult does cannot be conceptualized as a continuous process of merely adding additional strategies. Admittedly, the available data yield to other interpretations. But the possibility of this kind of discontinuity is an intriguing one, one that should be pursued further.

In the original proposal for this grant we outlined a number of possible experiments capitalizing on the same kind of optional cue deletion incorporated in the two relative clause experiments. These further experiments were intended to obtain information about the kinds of cues used when the optional cues were not present, i.e., information about alternative comprehension strategies. We noted in the proposal, for example, that the possibility of deleting the relative pronoun from a relative clause was constrained by the internal structure of the clause itself. Thus, while object relatives can apparently be reduced without exception, subject relatives cannot. Deleting the relative pronoun from (10) yields the ungrammatical (11). The reason for this constraint is obvious, for upon hearing (11) there is a tendency to treat liked as the main verb of the main clause, i.e., to be misled by the absence of cues indicating that liked is the verb of an embedded relative clause. It appears that subject relative clauses can be reduced in this way only when they contain non-finite verbs. Thus, in (12) the deletion is acceptable, for deletion removes not only the relative pronoun but also the auxiliary verb, leaving only the verb, wearing. And this, marked with the progressive -ing cannot be mistaken for a main clause main verb.

(10) The boy who liked the girl moved to Chicago.

(11) *The boy liked the girl moved to Chicago.

(12) The man (who was) wearing a gray hat entered the bar.

It appears, however, that the rules governing deletion contain a loophole. As Bever (1970) has noted, object relative clauses can be reduced even when they have been passivized, yielding sentences that are, in some cases, extremely difficult to understand, e.g., (13). The obvious problem is that raced tends to be interpreted as an active, intransitive verb with the horse as its subject rather than as a passive, transitive verb with the horse as its direct object and a deleted subject.

(13) The horse raced past the barn fell.

As we noted in the proposal, this difficulty does not appear to arise for all reduced passive relative clauses, for the difficulty comes from the identity of the past participle form of verbs like race with the simple past tense form of the same verbs. And there are some irregular verbs that do not suffer this identity of forms. Give, for example, yields given for its past participle and gave for its simple past.

Intuitively, the difficulty that arises with sentence like (13) does not appear to arise for sentences like (14).

(14) The boy given the magazine dropped it in a puddle and ruined it.

If this is the case, then it implies that information about the possible surface forms of a verb is retrieved and used sufficiently early in comprehension processing to allow the hearer to avoid being misled by the appearance of an NVN sequence (boy-given-magazine) into interpreting given as the active verb of the sentence's main clause. Thus, we proposed to compare the comprehension of sentences like (14) with that of ones like (15),

(15) The boy handed the magazine dropped it in a puddle and ruined it.

predicting that the latter would be considerably more difficult than the former even when difficulty was assessed with the phoneme monitor task on a monitoring target as early as the /m/ in magazine.

To make a long story short, the experiment has never been performed. The problem has been that we have been unable to find a sufficient number of pairs of verbs with the appropriate characteristics. This may seem surprising at first, but it becomes less so upon reflecting on the characteristics such verbs must have. First, there are relatively few verbs, even among irregulars, that have different past tense and past participle forms. For example, bring, although irregular, uses brought in both cases. In addition, usable irregular verbs tend to be frequently occurring ones; and since there is some evidence that monitoring latency is affected by frequency (e.g., Foss, 1969; Cairns & Foss, 1971), this argues that the regular verbs with which they are matched should be of equally high frequency. This, of course, considerably limits the selection of regular verbs, and we found that it limited it to the extent that we were unable to find very many regular-irregular pairs that were semantically sufficiently similar as to be equally appropriate in the same sentence frame.

Similar problems have plagued other experiments we had hoped to perform to test the use of other possible cues. For example, we had hoped to perform an experiment similar to the one just described in order to determine whether information about possible selectional restrictions between nouns and verbs might be used to avoid comprehension difficulty. That is, in Bever's kind of sentence (13) and in ours (14) the first nouns are animates and, as such, could logically be the subjects of the verbs that follow them. In general, it appears that the selectional restrictions between the subject and verb of a clause are more restrictive than those between the verb and the direct object. That is, there are many verbs that allow either animate or inanimate nouns as objects but only animate nouns as subjects. We had hoped that by finding such verbs we could construct sentences in which the verbs occurred in reduced passive relative clauses that differed in whether they had animate or inanimate direct objects, e.g., (16) and (17). The animate noun (astronauts) in (16) could be either the

(16) The astronauts raced to the NASA laboratory were found to have been exposed to high doses of radiation.

(17) The lunar rocks raced to the NASA laboratory were found to have been exposed to high doses of radiation.

subject or the object of *raced*, an ambiguity that is resolved only considerably later in the sentence. The inanimate noun (*lunar rocks*) in (17) could not be the subject of *raced*, and hence, (17) contains a constraint on the relationship between the first noun and the first verb that (16) does not contain. Thus, by comparing the comprehension difficulty of sentences like (16) with that of ones like (17) we had hoped to determine whether (and when) information about noun-verb selectional restrictions was used.

Again, our attempts to construct appropriate materials were beset with difficulties like those described earlier. In the course of trying, it became apparent that we needed to be concerned not only with the selectional restrictions between a verb and the subjects and objects it could take on its transitive reading but also with whether the verb also had an intransitive reading and, if so, with the selectional restrictions on its subject on that reading. The reason becomes apparent on inspection of sentence (16) and (17), for it is apparent that it is the intransitive reading of *raced* that (16) appears to allow but which is precluded in (17). Unfortunately, a great many verbs that require animate subjects in their transitive reading allow inanimate subject in their intransitive reading. In fact, all those verbs that linguists consider "causative" (e.g., *close*) have this characteristic; and this class appears to nearly exhaust the class of verbs having both transitive and intransitive readings. So, once again, we were unable to construct a sufficient number of appropriate sentence pairs to make the experiment possible.

Our difficulties in constructing materials for these experiments point to a very general difficulty in trying to explore the nature of comprehension strategies. We have been trying to follow the strategy of working small changes on sentences and observing their effects on comprehension difficulty. But it appears that all too often such small changes result in the sentences changing in meaning or in their becoming blatently ungrammatical (or, alternatively, impossible to construct while retaining grammaticality). That the language itself prevents such manipulations in perhaps valuable, for it prevents the occurrence of ambiguities and uncertainties that would result in miscomprehension or in extreme comprehension difficulty. But while making the task of understanding that much easier for the hearer, it also prevents us from exploiting such manipulations as a means of discovering how comprehension occurs.

III. Research on Ambiguity.

There are a number of problems concerning the effects of ambiguity on the process of comprehension which we had initially hoped to explore. Primary among these was the question of whether structural ambiguities (either surface structure or deep structure) have effects on on-line comprehension processing similar to those that have been found for lexical ambiguities. A number of experiments have found that phoneme monitoring latency immediately following the occurrence of an ambiguous word is longer than following a synonymous, unambiguous word (see Appendix III for a review of this research). Generally, this effect has been interpreted as indicating that when an ambiguous word is encountered its multiple semantic readings are activated or retrieved from long-term storage and are retained for some period of time before a decision is reached as to which reading is appropriate for the sentence being processed. A recent experiment by Cairns and Kamerman (personal communication) suggests that this effect is relatively short-lived. They found the usual ambiguity monitoring effect when the monitoring target was in the word immediately following the ambiguity, but the difference between the ambiguous and control words had disappeared when the monitoring target was moved two words further from the ambiguity.

This short-lived lexical ambiguity effect is consonant with results found in a sentence verification paradigm (Foss, Bever & Silver, 1966; Cairns, 1970, 1973). In these experiments subjects were presented with a sentence containing either a lexical or a structural ambiguity. Each sentence was followed by a picture (or a second sentence) that was congruent with one or the other reading of the ambiguity. The subjects' task was to say as quickly as possible whether the sentence and the picture (or the second sentence) were consistent. For lexical ambiguities, judgment latencies were longer when the picture (or sentence) matched the less likely reading of the ambiguity than when it matched the more likely reading, suggesting that by the time the comparison between sentence and picture occurred the less likely reading of the ambiguity was no longer available.

These same sentence verification experiments suggest, however, that there is a difference between the processing of lexical ambiguities and that of structural ambiguities, for the verification latency results for structural ambiguities reveal no assymetry between the less and more likely readings of the ambiguities. Given the assumptions indicated above in the account of the verification results for lexical ambiguities, the results for structural ambiguities suggest that the multiple possible structures are retained far longer than the multiple possible readings for ambiguous words.

This same asymmetry in the effects for lexical and structural ambiguities appears in sentence completion experiments in which the subjects are given sentence fragments and asked to complete them as quickly as possible (MacKay, 1966; Bever, Garrett & Hurtig, 1973). Thus there appears to be considerable reason for believing that lexical and structural ambiguities are processed differently. The question remains, however, of what the nature of the difference is. The existing verification and completion results suggest that the multiple readings for a structural ambiguity remain available longer than is the case for a lexical ambiguity. But do the effects arise in the same way? The phoneme monitoring results for lexical ambiguities indicate that on-line processing load increases no later than the beginning of the word following the ambiguous word. But when in processing does a structural ambiguity begin to affect comprehension processing?

The question has no simple, logical solution. For while it is possible to argue that the existence of an ambiguous word could be noticed as soon as the word's phonological representation was used to retrieve information from the internal lexicon, no such argument can be made for structural ambiguities. That is, when in the course of processing a sentence from left-to-right does it become apparent that more than one structure is possible, as in (18)? One could perhaps argue that the

(18) The shooting of the hunters ...

possibility of an ambiguity arises as soon as shooting is processed, for the existence of a gerundive nominal implies a reduced, underlying sentence. And since shoot has both transitive and intransitive readings, whatever noun phrase follows the gerund could be either the subject of the intransitive reading or the direct object of the transitive reading. On the other hand, it could equally well be argued that the ambiguity does not really arise until after the phrase of the hunters is processed. For until it becomes apparent whether the head noun of the prepositional phrase is human or non-human, the possibility of the multiple readings does not arise. That is, if elephants were substituted for hunters, it would have to be the case that shoot has its transitive reading, with elephants as object rather than subject (i.e., someone shot the elephants).

The problem is just that structural ambiguities, unlike lexical ambiguities, are not localized. And the question of when the possibility of multiple readings for a structural ambiguity is first considered becomes an interesting question about when the comprehender constructs hypotheses about the structural interpretation of an input sentence. If, as Fodor, Garrett and Bever (1968) have suggested, multiple structural hypotheses are formulated as early as possible (at least as soon as a verb is encountered), then the increased processing load that should be associated with a structural ambiguity should appear early. But just how early remains an open question, a question about when the possibility of the existence of a structural ambiguity first becomes evident.

Actually, the evidence that structural ambiguity has any effect on on-line processing load is quite weak. The only experiment that has raised the question is one by Foss (1970) who presented both lexically and structurally ambiguous sentences, together with appropriate unambiguous control sentences, in a phoneme monitoring experiment. The monitoring target for the structural ambiguities occurred after the entire ambiguous portion of the sentence was complete. Thus, for an ambiguity of the sort in (18) the monitoring target would have occurred a word or two after hunters. Foss reported that for both kinds of ambiguities, taken together, monitoring latency was longer than for the control sentences. But since the ambiguity-control comparison was not reported separately for the lexical and structural ambiguities, it is impossible to determine whether the entire effect was attributable to the lexical ambiguities or whether the structural ambiguities revealed the same effect.

Thus the qeustion about on-line processing effects of structural ambiguities is in part a question about whether such an effect exists at all. But equally important, it is also a question about where such an effect occurs. In considering how to approach these questions, we concluded that it would be unwise to start by simply performing an equivalent of the structural ambiguity portion of Foss's experiment. For even if structural ambiguity increased on-line processing load somewhere in the course of processing there is clearly no guarantee that the effect would appear at the point at which monitoring latency was measured, wherever that point might be. Thus it seemed that an entire series of experiments was required, each locating the monitoring target at a different point with respect to the location of the structural ambiguity. That this would require a series of experiments rather than just a single experiment, using the same set of sentences in a number of conditions that differed only in the location of the monitoring target, results from the characteristics of the phoneme monitor task itself.

The problem is, essentially, that the monitoring target is itself an integral part of the sentence. An efficient structural ambiguity-control comparison thus requires that the two sentences of such a pair be identical at the target point, i.e., have the same initial phoneme occurring in the same word. This is a difficult enough condition to meet when only a single target location is involved, and becomes even more difficult when many different target locations are desired. This difficulty is increased still further by the fact that the word-initial phoneme that serves as a monitoring target must be unique within the sentence. That is, if we wished to monitor on the /h/ in hunters in (18), /h/ could not occur in word-initial position elsewhere in the sentence. This condition clearly becomes extremely difficult to meet when many different monitoring targets are required within the same sentence. The limit on this becomes a condition in which each word in the sentence must begin with a different phoneme.

Because of the difficulty of doing this, it seemed to us that a better solution would be to seek a different on-line comprehension task that would yield the same comprehension difficulty effects as the phoneme monitor task but that would be more flexible. Presumably, the phoneme monitoring task is a split-attention task, attention being divided between attempting to understand the sentence being heard and attempting to identify the specified target phoneme. The sought-for task thus should also involve splitting attention between sentence understanding and some second task, where that task would involve responding to some stimulus that, unlike a phoneme monitoring target, was not an integral part of the sentence itself. Such independence of stimulus and sentence would allow the stimulus to be made co-incident with any part of the sentence desired.

Foss (personal communication) has attempted to develop one such task, a light-monitoring task. Essentially, while a subject is listening to a sentence (and trying to understand it), a brief light flash occurs to which the subject is to respond as rapidly as possible. Foss has performed several experiments using this task to determine whether it yields the same elevated latency following a lexical ambiguity as does the phoneme monitoring task. The results, unfortunately, have been uniformly disappointing. In no case has he found a significant latency difference using light monitoring between lexical ambiguities and their controls, even when using sentences and monitor target locations that yield such a difference for phoneme monitoring. Clearly, light monitoring does not seem a reasonable candidate.

When we first heard of Foss's results, it seemed to us that a possible reason for the ineffectiveness of the light monitoring task was the fact that the stimulus for which the response latency was being measured was not itself a linguistic stimulus. That is, the processing of a nonlinguistic visual stimulus may simply not engage the same processors as the sentence, an auditory linguistic stimulus. In addition, it seems likely that the processing load (of whatever sort) involved in responding to the presence of a light is considerably less than that involved in detecting and responding to the presence of a specified phoneme.

For these reasons, we felt we might have a better chance of success if the stimulus to be responded to were a linguistic one. So we turned to a variation of a task that has been used for a number of years for assessing some aspects of comprehension processing--the probe latency task. In its usual version this task involves presenting a sentence, followed at a short interval by a word. The subject's task is to decide whether the probe word occurred in the sentence. Since the probe response is a choice response, the task necessarily involves presenting sentence-probe pairs where the probe did occur in the sentence (IN probes) and ones where the probe did not occur (OUT probes). Attention has usually been focused on IN probes, and a number of experiments have examined latency to respond to IN probes as a function of the syntactic or semantic role the probe word played in the sentence.

Recently, however, Caplan (1971, 1972) has introduced a variation of the probe latency task which seemed to have some promise as a measure of on-line comprehension processing load. In Caplan's experiments, the probe occurred during rather than after the sentence, thus permitting studying probe latency as a function of where in the sentence the probe occurred as well as of the character of the probe itself.

In one experiment Caplan found that probe latency for IN probes was less when the probe was presented during the same clause as the word probed than when the probe was presented during a later clause, even when the distance between the probe and the word probed was constant. In another experiment Caplan studied probe latency following ambiguous and unambiguous words. While the results of this experiment were far from conclusive, they seemed to suggest that probe latency might yield results parallel to phoneme monitoring latency.

In addition, it seemed possible that latency to respond to OUT probes might furnish us with the kind of flexible measure we were seeking for examining structural ambiguity effects. One virtue of the probe task is that the point at which the probe is presented is independent of the content of the sentence, i.e., the probe can be presented conincident with any word in the sentence, early or late. In addition, since an OUT probe is a word that does not occur anywhere in the sentence there is no confounding of the syntactic or semantic role the word probed played in the sentence with the location at which it occurs.

But before attempting to use the on-line probe task to evaluate structural ambiguity effects we thought it wise to determine whether we could obtain some of the same effects with this task that we knew we could obtain with the phoneme monitoring task. Our first attempt in this direction was a replication of one of the relative clause experiments described in Appendix II. We used the sentences from the first of those experiments, the sentences which contained object relative clauses that were either self-embedded or right-branching and which could be presented with the relative pronoun either present or deleted. The probes for this experiment were all IN probes, the words probed being the head nouns of the NPs immediately preceding the relative clauses (i.e., the nouns that were coreferential with the relative pronouns and served as direct objects of the relative clauses). The probes for these sentences occurred at the same points in the sentences as had the phoneme monitoring targets. Conincident with what had been the target phoneme in the earlier experiment was an inaudible tone which served to open the shutter on a projection tachistoscope, presenting the relevant probe word. The subject's task was to respond as rapidly as possible by pressing one of two buttons to indicate whether or not the probe word had occurred in the sentence prior to its presentation as a probe. To make the experiment as nearly parallel to the earlier phoneme monitoring experiment, the subject was also required to paraphrase each sentence as soon as its presentation was completed.

To mask the purpose of the experiment, filler sentences were included that also contained IN probes which were not nouns and which occurred at a variety of locations in the filler sentences. There was, in addition, an equal number of sentences with OUT probes.

The results for the probe task were not exactly what we had hoped they would be. The latency effect for relative clause location (self-embedded vs. right-branching) was significant. Here, as in the phoneme monitoring data, responding was slower to right-branching object relative clauses than to the same clauses when they were self-embedded. We suspect, however, that the reason for this difference may not be the same as that for the similar difference in the monitoring experiment. Probes in right-branching clauses necessarily occurred later in their respective sentences than the probes in the self-embedded sentences. Since the probe task requires searching through that part of the sentence already processed to determine if the probe word had occurred, it should be the case that a probe later in a sentence should yield a longer latency than one earlier in the sentence simply because later there is more of the sentence to be searched. Thus this kind of early-late comparison is probably not one for which the on-line probe task is suitable.*

The probe latency difference for reduced and unreduced relative clauses was, unlike that obtained in the monitoring experiment, far from significant. The mean latency difference was in the expected direction; so out of curiousity we estimated, using the means and variances obtained from 40 subjects, the number of subjects necessary for a mean difference of the obtained magnitude to be significant. The number was slightly in excess of 3,000 per condition, suggesting that at best the on-line probe task is far less sensitive than the monitoring task.

The results of the experiment were hardly encouraging. One expected difference failed to materialize; the other did materialize, but probably for conceptually uninteresting reasons. Undaunted by this, we performed a second on-line probe experiment, searching for the lexical ambiguity processing effect that has been found so consistently with the monitoring task. The sentences here were taken from an unpublished phoneme monitor experiment by Foss and Hassler. Each of the test sentences contained either an ambiguous noun or an unambiguous control noun. In addition, half the time the relevant noun was preceded by a related adjective which might predisambiguate the ambiguous noun. The four variations for one sentence are presented in (19). Since these sentences had previously been

(19) a. The man bought the stock during... (Ambiguous, neutral context).

b. The rancher bought the stock during... (Ambiguous, biased context).

c. The man bought the cattle during... (Unambiguous, neutral context).

d. The rancher bought the cattle during... (Unambiguous, biased context).

used in a monitoring experiment we knew in advance that they yielded a significant ambiguity effect, i.e., longer latencies for the ambiguous than for the unambiguous versions. The effect for the biasing context in that monitoring experiment was in the expected direction, i.e., a reduction of the ambiguous-unambiguous difference, but the effect was not significant.

As noted earlier, we believed the "cleanest" form of the probe task was one in which the probes were OUT probes. So OUT probes were prepared for each sentence and were presented immediately after the ambiguous or unambiguous word (e.g., at the beginning of during for the sentences in (19)). The probe location was identical to the monitor target location. Additional OUT probe sentences, not involving ambiguity, were included, as were a substantial number of IN probe sentences.

---

*The phoneme monitoring task is apparently not greatly influenced by whether the monitoring target occurs early or late. Although no experiment has been performed to explore this directly, the evidence from several experiments indicates that there are no consistent early-late latency differences. Rather, if anything, monitoring latency may be more influenced by whether the target occurs early or late within a clause. But, again, the evidence is indirect.

At first glance the results seemed reasonably promising. The main effect for ambiguity was significant, being longer following the ambiguous words than following their unambiguous counterparts. So, in this respect, the probe results appeared to parallel those for phoneme monitoring. Also like the monitoring results was the fact that the context-ambiguity interaction was in the expected direction, though not significant.

Having discovered these effects, we were considerably heartened that at least the on-line probe task might give the right effects in its OUT probe version. Closer examination of the data dampened our enthusiasm a bit, for it appeared that the magnitude and even the direction of the ambiguity and context effects was influenced considerably by such things as the grammatical function of the ambiguous noun as well as being subject to other, possibly random, influences. In short, the fine-grained appearance of the probe data was neither terribly coherent nor terribly consistent with the appearance of the monitoring data. So, we remain cautiously optimistic, concluding that further experiments with the probe task, using materials where we knew what the monitoring effects were, would be advisable before attempting to use the probe task for new experiments like the structural ambiguity experiments.

We might note in passing that results obtained in other laboratories using the on-line probe task seem to be about as variable as our own. Garrett and Chodorow (personal communication) have performed a number of such experiments, most of them concerning the verb complexity problem, and have obtained uniformly negative results (i.e., lack of predicted differences). It seems possible, given our own verb complexity results with the phoneme monitoring task, that the difficulty here lies in the predictions rather than in the task. In addition, Cairns and Blank (personal communication) have used the on-line probe task to investigate deletion of both relative pronouns and complementizers, finding a significant effect for relative clause reduction (an effect inconsistent with our own probe data) but not for complementizer reduction. At this point, we are quite unsure whether or not the on-line probe task will repay the investment that has been made in it. In any event, we have not yet had the courage to attempt using it for studying whether and when structural ambiguity affects on-line processing load.

A second kind of research on ambiguity has met with rather more success. A question that has been of continuing concern to psycholinguists and to cognitive psychologists generally concerns the nature of the interaction between information about one word (or, more generally, one perceptual entity) and others that occur surrounding it. This we may refer to as the context problem, the question being, how does context influence the interpretation of the material for which it is the context?

One reasonable context in which to raise this question is that of ambiguous words, for here the existence of an ambiguous word creates a situation in which more than one interpretation is possible and where the selection of one or another interpretation may be influenced by the intrepretations of adjacent words and structures as well as by coincident non-linguistic events. In structuring the problem, it seems reasonable to suppose on the basis of the phoneme monitoring results on lexical ambiguity that upon encountering an ambiguous word in a sentence the multiple readings are retrieved from the mental lexicon. Since the evidence suggests that these readings remain available for at least a short period of time, it seems clear that if a context occurs after the ambiguous word and serves to indicate that only one reading of the ambiguity is appropriate there should be little difficulty using that context to select out one of the available readings of the ambiguity.

If a disambiguating context precedes the ambiguity, however, an interesting question arises. Does such a context serve to prevent readings of the ambiguity other than the one favored by the context from ever being accessed or activated? Or does such a prior context have its selective effects in the same way as in the case where the context follows the ambiguity, i.e., by selecting among readings all of which have been accessed? The former seems closer to the kind of view popularized by the Gestalt psychologists of context creating an "expectancy". But to date most research seems to favor the latter view. A series of phoneme monitoring experiments by Foss and his associates (reviewed in detail in Appendix III) have examined the ambiguity monitoring effect as a function of whether or not the ambiguous word is preceded by a disambiguating context. Fairly uniformly, the results have yielded non-significant reductions in the magnitude of the ambiguity-control difference in the presence of the context. If the prior context restricted access to only the relevant reading of the ambiguous word, it seems reasonable to suppose that monitoring latency following a predisambiguated ambiguity would be no longer than following an unambiguous word. Hence, Foss's results seem to favor the view that a prior context affects the interpretation of an ambiguous word only after all readings of the ambiguity have been retrieved.

A review of Foss's experiments and others on the same problem convinced us that each of the experiments contained methodological and/or materials problems that rendered their results less than definitive. So, we decided that another experiment was warranted. The experiment was performed by David Swinney as his dissertation research, and the dissertation is included as Appendix III to the present report. Since the research is presented in detail there, there is little to be gained by reviewing it here. Suffice it to note that, using very carefully constructed disambiguating contexts, Swinney found that a prior context could completely eliminate the monitoring ambiguity effect. Thus, at first glance, his results seem to support the view that a prior context really does predisambiguate an ambiguity, i.e., restrict the amount of information retrieved.

However, as we realized in reflecting on Swinney's results, there are compelling logical reasons for believing that the predisambiguation account cannot be correct and that the interpretation of his results is far more complicated than is at first apparent. So, the original question remains less than completely resolved. But the account which Swinney has developed of his results suggests some lines for further research which, though extremely difficult to pursue, seem conceptually interesting.

IV. Research on Idioms.

Our interest in ambiguity has led us into an additional line of research as well--research on the comprehension of idiomatic expressions. It has long been recognized that idioms are syntactically and/or semantically peculiar in that their meanings are not compositional functions of the meanings of their parts, combined according to the usual syntactic and semantic rules. For example, the idiomatic meaning of kick the bucket is not the kind of reading to be expected for a structure that is a verb phrase composed of a transitive verb and a noun phrase direct object; rather its idiomatic reading is more closely akin to that of the intransitive verb die.

Although the peculiarities of idioms have attracted the attention of a great many linguists (whose formulations are reviewed in Appendix IV), it appears that they have not been nearly so attractive to psycholinguists interested in comprehension processing. Other than our own research, there exists exactly one study of idiom comprehension (Bobrow & Bell, 1973). The basic problem is, of course, how does the hearer discover that he is dealing with an idiom rather than a "normal" sentence, and how, once this has been discovered, does he find the appropriate meaning for the idiom?

The problem is made more complex by the fact that there appear to be two kinds of idioms. For one kind (e.g., kick the bucket), the string that has an idiomatic reading has a normal, non-idiomatic reading as well. For the other (e.g., tripping the light fantastic), there appears to be no parallel, non-idiomatic reading, though it seems likely that at some point in the history of English there was. The first kind is essentially ambiguous between the literal and idiomatic readings while the second is unambiguous. In addition, the first has a straightforward grammatical reading while the second is, on any attempt to treat it literally, ungrammatical.

If seems to us reasonable to suppose that when a hearer processes a sentence that contains an idiom he attempts to process it in the normal way. Doing so, it is relatively easy to see how he might discover the existence of an ungrammatical idiom, one that has no literal counterpart. For here the hearer encounters a portion of the sentence for which normal comprehension processing routines fail. And we may suppose that attempting to find an idiomatic reading for the ungrammatical string is one of the special processing routines that are invoked when normal processing fails. In this sense, finding an idiomatic reading for an ungrammatical idiom probably shares many characteristics with other kinds of attempts to interpret ungrammatical sentences, the difference being that there is a "standard" interpretation available in the case of the idiom whereas other kinds of ungrammatical sentences probably require more constructive processes for their interpretation.

This kind of account of the discovery of the existence of an idiom poses a problem, however, for those idioms that also have literal interpretations. For here, normal processing routines allow construction of the literal, non-idiomatic reading. And unless those same routines uncover the existence of the idiomatic reading, there is nothing else to signal that such a reading exists. One might argue that the existence of the idiom would be signalled by the fact that the literal reading of, say, *kicked the bucket* would be inconsistent with the rest of the speaker's discourse and the context in which it was uttered and, hence, that the existence of the idiom would be detectable from the lack of fit of the sentence and its surroundings. But while this is no doubt possible, it clearly cannot account for most cases of discovering the idiomatic reading of an ambiguous idiom. Such idioms are clearly recognized in situations where the sentence occurs in isolation. Thus the idiomatic reading must be discoverable during normal comprehension processing. And in this respect, ambiguous idioms appear to be like other kinds of ambiguities.

We have performed a series of experiments on the comprehension of idioms, guided in part by kind of general thinking just outlined. The experiments were the dissertation research of Linda Brannon and are reported in detail in Appendix IV. As becomes clear there, our research has only barely scratched the surface of the problems involved in understanding how idioms are discovered and understood. Here, as elsewhere, the phenomena are far more complex than we initially anticipated, but we have been able to formulate an account of at least parts of what is involved, an account that suggests some interesting ways in which idioms are related to other ambiguities. Since both the research and the directions in which it seems to point are discussed in detail in Appendix IV, we will not repeat that discussion here.

V. Coda.

We would like to be able to title this section of the report "Conclusions" and to indicate in it a coherent view of the processes involved in comprehending sentences. But the current state of affairs is far too chaotic to permit that. Our research over the past two years has led to no great unifying principles of comprehension processing not to any radically new views of that processing. In some sense, the major insight occurred some years ago, the insight that comprehension processing is a heuristic process not closely linked to the rules of a transformational grammar. But given that insight, there remains a great deal of research and theorizing necessary to pick apart the details of how such processing proceeds. We have, we believe, made some progress in that direction, discovering a few of the details and gaining some insight into how some of those details fit together. But it is far too soon for any major synthesis to emerge. We anticipate, with considerable reluctance, many more years of the kinds of research we have described in this report before a coherent, detailed account of comprehension processing emerges.

References

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.) Cognition and the development of language. New York: Wiley, 1970. Pp. 279-362.

Bever, T. G., Garrett, M. F., & Hurtig, R. The interaction of perceptual processes and ambiguous sentences. Memory & Cognition, 1973, 1, 277-286.

Bobrow, S. A., & Bell, S. M. On catching on to idiomatic expressions. Memory & Cognition, 1973, 1, 343-346.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Brown, H. D. Children's comprehension of relativized English sentences. Child Development, 1971, 42, 1923-1936.

Cairns, H. S. Ambiguous sentence processing. Unpublished Ph.D. dissertation, The University of Texas at Austin, 1970.

Cairns, H. S. Effects of bias on processing and reprocessing of lexically ambiguous sentences. Journal of Experimental Psychology, 1973, 97, 337-343.

Cairns, H. S., & Foss, D. J. Falsification of the hypothesis that word frequency is a unified variable in sentence processing. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 41-43.

Caplan, D. N. Probe tests and sentence perception. Unpublished Ph.D. dissertation, M.I.T., 1971.

Caplan, D. Clause boundaries and recognition latencies for words in sentences. Perception and Psychophysics, 1972, 12, 73-76.

Clark, H. H. Comprehension and the given-new contract. Paper presented at the conference on "The role of grammar in interdisciplinary linguistic research," University of Bielefeld, Bielefeld, Germany, December 11-15, 1973.

DeVilliers, J. G., & DeVilliers, P. A. Development of the use of word order in comprehension. Journal of Psycholinguistic Research, 1973, 2, 331-341.

Fillmore, C. J. The case for case. In E. Bach & R. T. Harms (eds.) Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968. Pp. 1-90.

Fodor, J. A., & Garrett, M. Some syntactic determinants of sentential complexity. Perception and Psychophysics, 1967, 2, 289-296.

Fodor, J. A., Garrett, M., & Bever, T. G. Some syntactic determinants of sentential complexity, II: Verb structure. Perception and Psychophysics, 1968, 3, 453-461.

Foss, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 457-462.

Foss, D. J. Some effects of ambiguity upon sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 699-706.

Foss, D. J., Bever, T. G., & Silver, M. The comprehension and verification of ambiguous sentences. Perception and Psychophysics, 1968, 4, 304-306.

Hakes, D. T. Does verb structure affect sentence comprehension? Perception and Psychophysics, 1971, 10, 229-232.

Hakes, D. T. On understanding sentences: In search of a theory of sentence comprehension. Final Report, Grant OEG-0-70-3450(508), U.S. Office of Education, September, 1972. (a)

Hakes, D. T. Effects of reducting complement constructions on sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 278-286. (b)

Hakes, D. T., & Cairns, H. S. Sentence comprehension and relative pronouns. Perception and Psychophysics, 1970, 8, 5-8.

Hakes, D. T., & Foss, D. J. Decision processes during sentence comprehension: Effects of surface structure reconsidered. Perception and Psychophysics, 1970, 8, 413-416.

Healy, A. F., & Miller, G. A. The verb as a main determinant of sentence meaning. Psychonomic Science, 1970, 20, 372.

Healy, A. F., & Miller, G. A. The relative contribution of nouns and verbs to sentence acceptability and comprehensibility. Psychonomic Science, 1971, 24, 94-96.

Katz, J. J. Semantic theory. New York: Harper & Row, 1972.

MacKay, D. G. To end ambiguous sentences. Perception and Psychophysics, 1966, 1, 426-436.

Miller, G. A., & Chomsky, N. Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.) Handbook of mathematical psychology, Vol. II. New York: Wiley, 1963. Pp. 419-492.

Miller, G. A., & Isard, S. Free recall of self embedded English sentences. Information and Control, 1964, 7, 292-303.

Perfetti, C. A. Psychosemantics: Some cognitive aspects of structural meaning. *Psychological Bulletin*, 1972, 78, 241-259.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 1967, 2, 437-442. (a)

Sachs, J. S. Recognition of semantic, syntactic and lexical changes in sentences. Paper presented at the Psychonomic Society meetings, Chicago, Oct., 1967. (b)

Sheldon, A. The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 272-281.

Sinclair, H., & Bronckart, J. P. S.V.O. A linguistic universal? A study in developmental psycholinguistics. *Journal of Experimental Child Psychology*, 1972, 14, 329-348.

Slobin, D. I. Cognitive prerequisites for the development of grammar. In C.A. Ferguson & D.I. Slobin (Eds.) *Studies of child language development*. New York: Holt, Rinehart & Winston, 1973. Pp. 175-208.

Smith, M. D. Relative clause formation between 29-36 months: A preliminary report. *Papers and Reports on Child Language Development*, 1974, #8, 104-110.

Winograd, T. Understanding natural language. *Cognitive Psychology*, 1972, 3, 1-191.

APPENDIX I

Verb Complexity and Sentence Comprehension: Further Data and Reflections

David T. Hakes

Verb Complexity and Sentence Comprehension: Further Data and Reflections[1]

David T. Hakes

The University of Texas at Austin

Fodor, Garrett and Bever (1968; see also Bever, 1970) have proposed a theory of the processing involved in comprehending sentences which incorporates a process of actively "projecting" candidate syntactic structures for a sentence. They suggest that while the hearer is listening to a sentence he is formulating hypotheses, based upon cues contained in the sentence's superficial form, as to the structure underlying both the parts of the sentence already heard and parts yet to come. Specifically, they argue that stored lexical information about the predicate structures that could be associated with a given verb is retrieved when the verb itself is processed and is used to formulate a set of hypotheses about the predicate structure actually realized in the sentence being processed. Thus, for example, if the verb in some clause of a sentence has only an intransitive reading, the hearer hypothesizes that the material intervening between that verb and the next clause boundary contains either nothing or one or more adverbials (including prepositional phrases); he does not consider the possibility that the predicate contains a noun phrase (NP) or that it contains an embedded sentential complement construction. If the verb allows both intransitive and transitive readings, then in addition to the hypotheses of nothing and of adverbials the hearer must also consider the hypothesis that the predicate contains a NP.

In general, the theory yields the prediction that the greater the number of possible predicate structures associated with a verb, the greater the number of hypotheses the hearer must project and test against the material actually contained in the sentence to determine the correct structure. Assuming that formulating and testing a hypothesis involves some processing effort, then the greater the number of possible structures the more difficult the sentence should be to comprehend. It is important to note that Fodor et al's theory claims that hypotheses about predicate structure are projected when the verb is encountered and before sufficient information about the predicate structure actually realized becomes available to remove the uncertainty, i.e., it is a theory of active, on-line comprehension processing.

Fodor et al reported two experiments designed to test this verb complexity hypothesis, both involving comparisons of sentences containing "pure" transitive verbs (and, hence, NP predicates for those verbs) with sentences that were identical except for containing complement verbs, i.e., verbs allowing at least NP or embedded sentential complement predicates. In the first experiment Ss were required to paraphrase such sentences; in the second they were required to reconstruct them from anagrams. Both experiments yielded results consistent with the hypothesis, e.g., paraphrasing was more accurate (and more rapid) when the sentences contained transitive verbs than when they contained complementizing verbs.

Hakes (1971) has pointed out that neither of these experiments provided a direct test of the claim that verb complexity affects on-line comprehension processing or even comprehension processing at all. Paraphrasing must reflect processes involved in the construction of the paraphrase, regardless of whether it also reflects processes involved in comprehension. Thus, it remains unclear whether the paraphrasing data reflected differences in comprehension difficulty and, even if they did, whether those differences could be attributed to differences in on-line processing or to some aspect of comprehension processing that occurred after the sentence had been completed. Similarly, it remains entirely unclear what relation there might be between the processes involved in constructing a sentence from a random array of words (i.e., an anagram) and those involved in comprehending a heard sentence.

In order to evaluate the verb complexity hypothesis more directly Hakes (1971) performed two experiments using sentences which, like those of the Fodor et al experiments, differed only in whether they contained a transitive or a complementizing verb occurring with a NP predicate. In both experiments Ss were required to paraphrase the sentences and, in addition, to perform a phoneme monitoring task involving listening for a specified word-initial phoneme and responding manually as rapidly as possible when it was detected. A number of experiments have suggested that this task is a sensitive measure of on-line processing load during sentence comprehension, though as yet no comprehensive account is available of what aspects of comprehension processing are reflected in monitoring latency or how (see Hakes, 1972b, for a review of much of this research). For one experiment the target phoneme was either the initial phoneme of the verb itself or that of the head noun of that verb's object NP; for the other experiment it was always the initial phoneme of a word in the object NP (i.e., always the second word after the verb).

The results of Hakes' experiments may be stated quite simply. The paraphrasing task yielded a significant verb complexity effect for one experiment and a non-significant effect in the predicted direction for the other. For neither experiment did the phoneme monitoring task yield the predicted effect. Taken together, the results for the several experiments on verb complexity lend little support to the hypothesis that verb complexity increases processing load while the sentence is being processed and, more particularly, before the predicate for the relevant verb has been processed.

There is, however, some reason to believe that the effects obtained in the paraphrasing task do reflect comprehension processing effects, i.e., that they are not entirely attributable to post-comprehension differences in the difficulty, e.g., of constructing paraphrases. While the verb complexity experiments themselves could yield to the latter interpretation, other paraphrasing experiments cannot. For example, Hakes (1972a) presented paraphrasing (and phoneme monitoring) data from two experiments comparing sentences all of which contained complementizing verbs and which differed only in whether they contained the optional complementizer "that" as a cue to the existence of an embedded complement construction. For one experiment the sentences containing the complementizer were paraphrased significantly more accurately than those which did not; for the other there was a non-significant difference in the same direction. Since the underlying structure and meaning of the sentences were identical whether the complementizer was present or not, the paraphrasing effects must be attributable to differences in the difficulty (or likelihood)of recovering that structure and meaning correctly, i. e., to differences in comprehension difficulty. The same argument applies to the paraphrasing results of several experiments comparing sentences containing full relative clauses with ones containing reduced versions of the same relative clauses (Fodor & Garrett, 1967; Hakes & Cairns, 1970; Hakes & Foss, 1970). Arguing from like effects to like causes, it follows that the paraphrasing effects obtained in the verb complexity experiments are also attributable to differences in comprehension difficulty.

Considering the data from the paraphrasing and phoneme monitoring experiments, the following picture emerges. Verb complexity does appear to affect processing difficulty during comprehension, but it does not appear to affect it at the point in processing suggested by Fodor et al. What, then, is the locus and, for that matter, the nature of the effect?

Intuitively, it seems implausible that information about a verb's predicate structure possibilities should be used in quite as "active" a way as Fodor et al suggest. Why should the hearer project the set of possible structural hypotheses when first encountering the verb? Generally, one need not wait for long after the occurrence of a verb for there to be sufficient information available to uniquely determine the nature of the predicate. That is, verbs are generally not separated from the clause boundary marking the end of the predicate by very much material that isn't relevant to the nature of the predicate.

Consider, for example, a verb that allows either NP or sentential predicate structures. If, in fact, it is the transitive structure (NP) that is realized, then the structure is generally uniquely determined shortly after the end of the sequence analyzable as NP. (We will consider an exception to this later.) To be sure, that NP may have a complex internal structure, as for example, when it contains complex modifiers. But the point is that there seems to be no intuitively plausible reason to try to anticipate which structure will be realized when generally which one is actually realized is soon apparent.

If, in fact, it is the case that a decision about the structure of a verb's predicate cannot be made until the end of that predicate, i.e., until the following clause boundary, then an alternative account of the locus of the verb complexity effect becomes apparent.[1] We suggest that when the hearer determines that a clause has concluded he attempts to integrate the verb with the rest of the predicate to assign a structure to it. If it is more difficult to assign a structure when the verb permits a number of different structures, then it is at this point, shortly after the objective occurrence of the clause boundary, that a verb complexity effect should manifest itself.

---

[1] See Garrett and Bever (in press) for a discussion of research suggesting that the ends of clauses are points at which considerable processing occurs, presumably processing involving the assignment of a structure to the preceding clause.

The first two experiments reported below were designed to test this hypothesis. Both used both the paraphrasing and phoneme monitoring tasks and the target phonemes for the latter task were located shortly after the end of the clause containing either a transitive or complementizing verb and a NP direct object. The third experiment was designed to test a variant of the clause boundary hypothesis. Since by the time the end of the clause has been reached the predicate structure is usually uniquely determined, it may be no more difficult in such cases to assign a given structure when the verb permits many structures than when it permits only that structure, i.e., it may be no more difficult to assign a transitive structure when the verb permits sentential predicates than when it does not if the rest of the predicate is unambiguously a simple NP (i.e., one not containing an embedded sentence). If this is the case, then an effect of verb complexity should be apparent only if the structure of the predicate is not uniquely determined by the end of the clause. This possibility was evaluated by constructing sentences having this character for use in a phoneme monitoring and paraphrasing experiment.

EXPERIMENT I

Method

Design and Materials. Twenty test sentences were constructed to serve as frames for either transitive or complementizing verbs. For each such sentence two verbs were selected, one a "pure" transitive and the other allowing both simple NP direct objects and NP object complement constructions. The verbs in each pair were matched for frequency of occurrence in the Thorndike-Lorge (1944) norms. A concerted effort was made to ensure that each sentence frame was sensible with either verb. The sentence frames were complex, containing 3-5 clauses. A variety of structural complexities occurred, including several kinds of subordinate clauses, complex modifiers and adverbial and prepositional phrases. For all sentence frames, the clause containing the verb of interest was a main, independent clause containing a simple NP direct object (i.e., one containing neither complex modifiers nor embedded sentences). This was followed by a co-ordinate independent clause, the boundary between these clauses being as clearly marked as possible.

Each test sentence was designed so that a word (the target word) early in the second of the co-ordinate clauses contained the phoneme to be monitored. This was either the word immediately following the co-ordinate conjuction or the second word following. The target phoneme did not occur in word-initial position elsewhere in the sentence.

Two sets of filler sentences were also constructed. One set of 20 sentences contained complementizing verbs followed by NP object complement constructions. These served to prevent Ss from developing the expectation that a complementizing verb would always be followed by a simple NP object.

The second set of filler sentences consisted of 24 complex sentences designed without any particular structural constraints. Sixteen of these contained target words in locations varying widely from that for the test sentences. The remaining eight contained no target word and were included to prevent Ss from expecting that such a word would always occur. There was thus a total of 64 sentences, 20 of which could be presented with either a transitive or a complementizing verb.

A single, semi-random order of the 64 sentences was prepared, and two tape recordings of the sentences in this order were made. On one tape, a subset of 10 test sentences contained transitive verbs, the other subset contained complementizing verbs. On the other tape, the assignment of verb types to test sentences was reversed. Tones of 1000 Hz were recorded on a separate track of the tape, coincident with the onset of each target phoneme, to be used to activate the timer used for recording monitoring latency.

Each sentence was preceded on the tape by an instruction indicating the target phoneme for that sentence. For the filler sentences that contained no target phoneme the instructions specified a phoneme that did not occur in word-initial position in that sentence. Instructions for the phoneme monitoring and paraphrasing tasks, as well as four practice sentences, were also recorded and dubbed onto the beginning of each tape.

To summarize, the design included two groups of Ss, defined by the two tapes, i.e., by the assignment of verb types to test sentences. Two experimenters each tested half the Ss in each group. Within Ss, sentence containing both transitive and complementizing verbs occurred. The design was thus a 2 x 2 x 2 mixed factorial.

Procedure. Forty students from the freshman-level introductory psychology course served as Ss. Participation in experiments was a course requirement, but the Ss were not required to serve in this particular experiment. All were native speakers of English, and none had previously served in any experiment involving sentence comprehension, phoneme monitoring or paraphrasing. Approximately equal numbers of females and males served, though Ss were not selected or assigned to groups on this basis.

Results and Discussion

Phoneme Monitoring. Individual monitoring latencies were converted to speed scores by reciprocal transformation. For each S, two mean speed scores were computed, one across the sentences in the transitive verb condition and the other across the sentences in the complementizing verb condition. All analyses were performed on these mean speed scores. The means of the mean speed scores, taken across groups and experimenters, are presented in Table 1. Monitoring appears to have been somewhat faster when the sentences contained transitive verbs than when they contained complementizing verbs.

Insert Table 1 about here

The mean speed scores were subjected to a Groups x Experimenters x Transitive vs. Complementizing Verb analysis of variance. No main effects or interactions reached significance, the main effect for Verb Type yielding $\underline{F}$ (1,36) = 1.91, $\underline{p} = .17$. Thus the verb complexity effect, though in the predicted direction, is not such as to inspire any confidence that it is a real effect.

The results of a sentence-by-sentence analysis, while not orthogonal to the results reported above, clearly confirm them. A Wilcoxen matched-pairs, signed-ranks test by sentence yielded a value that did not even approach significance. Thus it seems clear that there is not even a small effect that is consistent across sentences or subjects.

Paraphrasing. Each paraphrase response was transcribed for scoring. Each test sentence was divided into segments corresponding roughly to surface structure clauses, and the paraphrases were scored in terms of the proportion of a sentence's segments that were paraphrased accurately. To be counted as correct both the internal grammatical relations of a segment and its relations to the other segments had to be accurately represented in the paraphrase. Lexical substitution was liberally allowed, except that substitution of an indefinite for a definite was not accepted (e.g., someone for the blind student). Verbatim reproductions of the test sentence were not accepted.

The paraphrase responses were scored independently by two judges, each judge scoring all paraphrases of a given test sentence without regard to whether it had occurred with a transitive or a complementizing verb before proceeding to the paraphrases of another sentence. In addition, paraphrases of the sentences from the present experiment were intermixed during scoring with those from another experiment. The judges did not know during scoring whether a given paraphrase came from the transitive or complementizing verb condition or even from which experiment it came. Differences in the scores assigned by the two judges were resolved by discussion after all scoring was completed. (Further discussion of paraphrase scoring criteria and procedures may be found in Fodor et al, 1968; Hakes, 1971, 1972a).

Since the number of segments scored varied across test sentences, a S's accuracy score for any given sentence was taken as the proportion of the segments paraphrased accurately. Two mean accuracy scores were then computed for each S, one for sentences containing transitive verbs and the other for ones containing complementizing verbs. The means of these mean accuracy scores, taken across groups and experimenters, are presented in Table 1. It is apparent that the verb type conditions differ in the predicted direction; the test sentences were paraphrased considerably more accurately when they contained transitive verbs than when they contained complementizing verbs. This conclusion is borne out by the results of a Groups x Experimenters x Verb type analysis. The only main effect to reach significance is that for Verb Type, $\underline{F}$ (1,36) = 8.63, $\underline{p} = .006$. There was, however, a significant Groups x Verb Type interaction, $\underline{F}$ (1,36) = 15.79, $\underline{p} < .001$. Inspection of the data indicates that for one assignment of verb types to test sentences, the effect was strongly in the predicted direction while for the other assignment it was marginally in the opposite direction.

To determine whether the effect of verb type could be considered a meaningful one, the paraphase data were also subjected to a sentence-by-sentence analysis. A Wilcoxen matched-pairs, signed-ranks test indicated that the test sentences were consistently paraphrased more accurately ($\underline{p} < .001$) when they contained transitive verbs than when they contained complementizing verbs. Although this analysis is not orthogonal to that reported above, its outcome does strengthen the conclusion that sentences containing transitive verbs are consistently paraphrased more accurately than ones containing complex verbs used as transitives.

For the paraphrasing task, then, the record now stands at three successes, i.e., experiments that have found a significant verb complexity effect (Fodor et al, 1968; Hakes, 1971, Expt. 2; the present experiment), and one failure (Hakes, 1971, Expt. 1). Hakes (1972a) has argued on the basis of a large number of experimental results that the paraphrasing task is relatively insensitive to effects of the sort sought here, so it is perhaps not unreasonable that the data are not entirely consistent in showing a verb complexity effect. Further, if the argument advanced above is correct, then at least a major portion of the obtained effect must be attributed to a difference in the difficulty of processing sentences with simple and complex verbs during comprehension. The puzzle that remains is why such an effect should fail to reveal itself in the phoneme monitoring task, a task generally far more sensitive to such on-line processing effects than the paraphrasing task.

However, before attempting to account for the lack of a phoneme monitoring effect, it seemed wise to determine more certainly if there was no such effect observable shortly after the end of the clause. Recall that the monitoring results of Expt. 1 did yield a difference in the predicted direction, albeit one far too small to achieve statistical significance. For this reason, Experiment 2, essentially a replication of Experiment 1 with some modifications, was performed.

EXPERIMENT 2

Method

Design and Materials. The design and materials here were highly similar to those of Expt. 1. Sixteen test sentences, each allowing either a transitive or a complementizing verb in one of its main clauses, were constructed. Five of these had also been used in Expt. 1, while the remaining 11 were new. Sixteen filler sentences, having the same structural characteristics as the first set of fillers in Expt. 1, were also constructed, as were 32 filler sentences corresponding to the second set of Expt. 1 fillers. Fourteen of the latter were designed to occur with no target phoneme. In addition, the filler sentences were designed to be structurally more complex than their Expt. 1 counterparts.

The assignment of test sentences to groups (tapes) and all details of the construction of the stimulus tapes were identical to those of Expt. 1. Again, there were two groups, defined by the assignment of verb types to test sentences; two experimenters each tested half the Ss in each group; and within Ss, both verb types occurred.

Procedure. All details of the procedure were identical to those of Expt. 1.

Subjects. Forty Ss having the same general characteristics as those in Expt. 1 participated. Ten were assigned to each combination of groups and experimenters according to a random schedule.

Results and Discussion

Phoneme Monitoring. All monitoring latencies were again converted to speed scores, and mean speed scores for each S were computed across the sentences in each of the verb type conditions. The means of these mean speed scores, taken across groups and experimenters, are presented in Table 2. Unlike Expt. 1, monitoring here was slighly slower when the sentences contained transitive verbs than when they contained complementizing verbs.

The mean speed scores were subjected to a Groups x Experimenters x Verb Type analysis of variance. The only main effect to reach significance was that for Experimenters, F (1,36) = 4.133, p =.047. The effect for Verb Type was far from significance, F (1,36) = 0.305, p = .59. No interaction reached significance. Rather clearly, any comprehension difficulty associated with verb complexity is not reflected in monitoring latency measured immediately after the end of the clause containing the relevant verb.

Paraphrasing. The paraphrasing data were treated in a manner identical to that described for Expt. 1, being scored by the same two judges, using the same procedures and criteria. Each S's score was again converted to a proportion of a sentence's segments paraphrased accurately, and the means of these proportions were computed for each S across the sentences contributing to each verb type condition. The means of these proportions, computed across groups and experimenters, are presented in Table 2. Here, as in Expt. 1, the sentences were paraphrased more accurately when they contained transitive verbs; the difference, however, is condierably smaller.

The mean proportion accuracy scores were subjected to analysis of variance. The main effect for Verb Type was not significant, F (1,36) = 2.439, p = .123. No other main effect or interaction approached significance. Thus the verb complexity effect found in previous paraphrasing experiments is only hinted at here. Given the insensitivity of the task, this is perhaps not too surprising.

The question that remains is, where do we go from here? Considering all the paraphrasing data, it seems that verb complexity does affect the difficulty of sentence comprehension. Yet whatever such effect exists does not manifest itself in the phoneme monitoring task, either when sought in the predicate of the clause containing the relevant verb or when sought immediately after the end of that clause. The possibility remains that we have simply been looking in the wrong place for the effect, e.g., that processing is slower than we have assumed and, hence, the effect should be sought still further from the verb. But without some account motivating a search at some particular point, such a search would resemble the proverbial search for the needle.

Before abandoning verb complexity as an intractable problem, there is another possibility we wish to consider. Assume that the basic clause boundary account is correct; that immediately after locating the end of a clause the hearer assigns a structure to the contents of that clause. Now, it is generally the case that by the end of a clause a unique structure has been determined for that clause, i.e., the material in the predicate of the clause allows an unambiguous interpretation of the appropriate reading for the verb. Roughly, if the predicate contains only something analyzable as a simple NP, then the verb must have its transitive reading, regardless of its other possible readings. If such information is used in assigning a reading for the clause, then there is no particular reason that other readings of the verb should be considered at all or should contribute to the difficulty of structure assignment. Hence, there should generally be no greater difficulty associated with a complex verb than with a simple verb.

There is, however, an exception. Sentences can be constructed in which it remains unclear for an indefinite period what the appropriate predicate structure is. Consider, for example, sentence (1) which contains a complementizing verb, followed by a simple NP which, in turn, is followed by a non-restrictive relative clause. For such sentences it is not clear until after the end of the relative clause whether the NP is the direct object of the verb or the subject of an embedded complement, for the sentence could have concluded as in sentence (2).

(1) At his press conference, the ambassador explained the treaty, which he had negotiated secretly, and the press was amazed to learn about it.

(2) ... secretly, would result in increased trade with China.

If, as we have been assuming, the hearer were to attempt to assign structure after the treaty, he would for such cases, be faced with two options. First, he could defer the decision, pending further information. To do so should "cost" something in storage, at least on the assumption that storing relatively less analyzed material costs more than storing relatively more analyzed material. Second, he could "guess" at which structure was going to be correct, thus standing in danger of being wrong and having to change the decision after the end of the relative clause. That, too, should "cost" something, though later in processing than the first option.

Neither option is necessary for sentences like (3) or (4).

(3) ... the ambassador presented the treaty, which ...
(4) ... the ambassador explained that the treaty, which ...

Since (3) contains a verb not allowing complement constructions, the treaty is determined to be the direct object of presented at the clause boundary marking the beginning of the relative clause. For (4), although the verb allows either NP or sentential predicates, the complementizer that uniquely determines that there must be an embedded complement. Thus, we predict that sentences like either (1) or (2) should be more difficult to comprehend than ones like either (3) or (4). Experiment 3 was performed to test this version of the verb complexity hypothesis.

EXPERIMENT 3

Method

Design and Materials. Twenty-five sentence frames were constructed which, like sentences (1)-(4), could take either a transitive or a complementizing verb as a main verb. There were four versions of each test sentence. In one, the transitive verb was followed by a direct object NP on which was embedded a non-restrictive relative clause; this in turn was followed by a co-ordinate clause. The second version was identical to the first except that a complementizing verb was substituted for the transitive. The third version also contained the complementizing verb, followed by the complementizer that, followed by the same NP and relative clause as the first two versions; in this case, however, the relative clause was followed by the predicate of an embedded complement sentence (e.g., as in (2)). The fourth version was identical to the third, except that the complementizer that was omitted. For the first three versions, the monitoring target was located within the relative clause on either the second or third word following the relative pronoun (e.g., the n in negotiate in sentence (1)). For the fourth version and for a fifth, identical to the third, the monitoring target was located after the end of the relative clause, i.e., after it was unambiguously determined that the predicate structure included an embedded complement.

Thirty-five filler sentences were also constructed, 15 of which contained no monitoring target. The fillers exhibited a variety of complex constructions, and the monitoring targets for those sentences which had them were in locations unlike those involved in the test sentences.

The test sentences were randomly divided into 5 subsets of 5 each. Five stimulus tapes were prepared, differing only in the assignment of test sentences to the five conditions defined by the five versions of each sentence. The design thus included five groups, defined by the tapes. Two experimenters each tested half the Ss in each group. And within Ss there were five verb type conditions.

Procedure. All details of the instructions and procedure were identical to those of Expt. 1.

Subjects. Eighty Ss having the same general characteristics as those in Epxt. 1 participated. Eight were assigned to each combination of groups and experimenters according to a random schedule.

Results and Discussion

Phoneme Monitoring. All monitoring latencies were converted to speed scores; five mean speed scores were computed for each *S*, one for the 5 sentences in each condition. The means of these mean speed scores, taken across groups and experimenters, are presented in Table 3. Of the three conditions for which the monitoring target occurred within the relative clause (1,2, and 3), monitoring speed appears to have been slowest for condition 1 in which the verb was transitive. This, of course, lends no support whatsoever to the verb complexity hypothesis. The condition in which the complementizing verb was followed by the complementizer (Condition 3) appears to have yielded somewhat slower monitoring than the condition in which the verb was followed immediately by the NP (Condition 2), contrary to previous results (Hakes, 1972a). Finally, there appears to be little difference in speed as a function of whether the complementizer was present for sentences where the monitoring target occurred after the relative clause (Conditions 4 and 5).

Insert Table 3 about here

The mean speed scores were subjected to a Groups x Experimenters x Verb Type analysis of variance. The main effect for Verb Type was highly significant, *F* (4,280) = 3.882, $p < .01$. The only other effects to reach significance were Groups, *F* (4,70) = 2.757, $p < .05$, and the Groups x Verb Type interaction, *F* (16,280) = 4.314, $p < .01$. To explore the nature of the verb Type effect further a series of planned, paired comparisons was performed among Conditons 1, 2 and 3 and between Conditions 4 and 5. The only one of these to reach significance was that between Conditions 1 and 2, *F* (1,280) = 8.056, $p < .01$, indicating that monitoring was significantly faster when the sentences contained complementizing verbs followed by simple NP objects than when they contained transitive verbs followed by the same objects.

The interpretation of this difference between sentences with complementizing and transitive verbs is complicated by the significant Groups x Verb Type interaction, which suggests that the effect may not have been consistent in direction across test sentences. To determine whether this was the case, a sentence-by-sentence analysis was performed for those two conditions. A Wilcoxen matched-pairs, signed-ranks test indicated that monitoring was not consistently faster when the sentences contained complementizing verbs (*p* .05). Thus it would appear that the present experiment, like its predecessors, failed to yield any consistent difference in monitoring for sentences differing only in whether they contained a transitive verb or a complementizing verb occurring with a transitive reading.

Paraphrasing. The paraphrasing data were again treated as in Expt. 1. Five proportion accuracy scores were computed for each *S*, corresponding to the five verb type conditions. The means of these proportions are presented in Table 3. Here Conditions 3 and 4 represent essentially identical sentences. For both, the sentences contained complementizing verbs in complement constructions introduced by the complementizer *that*; they differed only in the location of the monitoring target.

It appears that paraphrasing was more accurate when the sentences contained transitive verbs than when they contained complementizing verbs with transitive readings (1 vs. 2). It also appears that paraphrasing was more accurate for sentences containing complement constructions when those constructions were introduced by the complementizer (4 vs. 5). Finally, there is some suggestion that paraphrasing may have been been more accurate when a complement verb was followed by a complement construction than when it was followed by a transitive construction (3 vs. 2), though this difference is quite small.

The mean proportion accuracy scores were subjected to a Groups x Experimenters X Verb Type analysis of variance. The main effect for Verb Type was highly significant, F (4,280) = 5.406, p .001. In addition, the main effect for Experimenters was also significant, F (1,70) = 3.919, p .05. This variable did not, however, enter into any significant interactions. The only significant interaction was that for Groups x Verb Type, F (16,280) = 10.677, p .001. A series of planned, paired comparisons was performed on the Verb Type conditions. The transitive-complementizing verb comparison (1 vs. 2) was highly significant, F (1,280) = 3.94, $p < .01$. The comparison between sentences containing complement constructions with and without the complementizer (4 vs. 5) was also highly significant, F (1,280) = 8.72, $p < .01$. The conditions in which a complementizing verb was followed by either a complement or a transitive construction (3 vs.2) did not differ, F (1,280) <1.

It appears that, unlike the monitoring data, the paraphrasing data repeat the results of earlier experiments. Not only do the present data reveal a significant verb complexity effect, they also replicate the complement reduction effect obtained earlier (Hakes, 1972a).

We should note in passing that the lack of a paraphrasing accuracy difference between sentences containing complement verbs with direct objects and ones containing complement verbs with complement constructions appears to be inconsistent with a finding reported by Holmes and Forster (1972). Using a task in which the words of a sentence are visually presented in serial order at a rapid rate, they found that sentences containing complement constructions were reported as accurately as simple (transitive) sentences. The latter were, however, reported more accurately than sentences containing complementizing verbs occurring in transitive constructions. However, in considering their findings, it should be noted that their comparison of complement construction sentences with transitive sentences involved three kinds of complements, one of which (*for-to* VP complements) was considerably easier than the other kinds of complements. Inspection of their results suggests that for *that*-NP complements (the kind used in the present experiments) performance was as bad as for sentences containing complementizing verbs in transitive constructions. Hence, the inconsistency is more apparent than real.

The puzzle with which we started remains. The post-sentential paraphrasing task indicates that sentences containing complex, complementizing verbs are more difficult than ones containing simple, transitive verbs. Further, the same effect obtains with quite a different post-sentential measure of difficulty, the serial visual presentation task (Holmes & Forster, 1972). But equally consistently, the on-line phoneme monitoring task has failed to reveal a verb complexity effect. That the problem is not just a problem with the monitoring task is suggested by the results of a series of unpublished experiments by Garrett and Chodorow (Garrett, personal communication). Using an on-line probe latency task, Garrett and Chodorow have consistently failed to find any effect of verb complexity.

All this would suggest, contrary to the argument suggested earlier, that whatever the effect of verb complexity is it is not an effect that occurs during the on-line comprehension processing of the sentence. If this is the case, then the argument for an active, on-line comprehension theory of the sort suggested by Fodor et al (1968) and by Bever (1970) is considerably weakened, at least to the extent that such theories give a central role of the testing of hypotheses about the possible structures associated with a sentence's verbs. Unfortunately, if this is the case, it also remains entirely unclear what is responsible for the complexity effect consistently found in post-sentential measures.

REFERENCES

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.) _Cognition and the development of language_. New York: Wiley, 1970, pp. 279-362.

Fodor, J. A., & Garrett, M. Some syntactic determinants of sentential complexity. _Perception & Psychophysics_, 1967, _2_, 289-296.

Fodor, J. A., Garrett, M., & Bever, T. G. Some syntactic determinants of sentential complexity, II: Verb structure. _Perception & Psychophysics_, 1968, _3_, 453-461.

Garrett, M., & Bever, T. G. The perceptual segmentation of sentences. In T. G. Bever & W. Weksel (Eds.) _The structure and psychology of language_. Cambridge, Mass.: M.I.T. Press (in press).

Hakes, D. T. Does verb structure affect sentence comprehension? _Perception & Psychophysics_, 1971, _10_, 229-232.

Hakes, D. T. Effects of reducing complement constructions on sentence comprehension. _Journal of Verbal Learning and Verbal Behavior_, 1972, _11_, 278-286. (a)

Hakes, D. T. On understanding sentences: In search of a theory of sentence comprehension. Final Report, Grant OEG-0-70-3450(508), U.S. Office of Education, September, 1972. (b)

Hakes, D. T., & Cairns, H. S. Sentence comprehension and relative pronouns. _Perception & Psychophysics_, 1970, _8_, 5-8.

Hakes, D. T., & Foss, D. J. Decision processes during sentence comprehension: Effects of surface structure reconsidered. _Perception & Psychophysics_, 1970, _8_ , 413-416.

Holmes, V. M., & Forster, K. I. Perceptual complexity and underlying sentence structure. _Journal of Verbal Learning and Verbal Behavior_, 1972, _11_, 148-156.

Thorndike, E. L.,& Lorge, I. _The teacher's word book of 30,000 words_. New York: Teachers College, Columbia University, 1944.

FOOTNOTES

[1] The research reported here was supported by Grants OEG-0-70-3450(508) and OEG-0-72-3941 from the U.S. Office of Education (now the National Institute of Education). I wish to thank Barbara Tabor, Thomas Harris and Jeff Lehmann for their assistance with the first two experiments and Jack Lumbley and John Mastenbrook for their assistance with the third experiment.

TABLE 1

Phoneme Monitoring Speed and Paraphrasing Accuracy:

Experiment 1

Verb Type	Speed	Accuracy
Transitive	4.71	.527
Complement	4.37	.468

TABLE 2

Phoneme Monitoring Speed and Paraphrasing Accuracy:

Experiment 2

Verb Type	Speed	Accuracy
Transitive	2.65	.487
Complement	2.70	.459

TABLE 3

Phoneme Monitoring Speed and Paraphrasing Accuracy:

Experiment 3

Verb Type	Predicate	Monitor	Speed	Accuracy
(1) Transitive	Direct Object	In rel. cl.	1.848	.552
(2) Complement	Direct Object	In rel. cl.	1.959	.516
(3) Complement	Complement (+that)	In rel. cl.	1.903	.526
(4) Complement	Complement (-that)	After rel. cl.	1.838	.468
(5) Complement	Complement (+that)	After rel. cl.	1.830	.522

APPENDIX II

Understanding Sentences with Relative Clauses

David T. Hakes, Judith S. Evans and Linda L. Brannon

Understanding Sentences with Relative Clauses

David T. Hakes, Judith S. Evans and Linda L. Brannon
Department of Psychology
The University of Texas at Austin
Austin, Texas 78712

Short Title: Understanding Sentences with Relative Clauses

Send Proofs to : Dr. David T. Hakes
Department of Psychology
The University of Texas at Austin
Austin, Texas 78712

Understanding Sentences with Relative Clauses

David T. Hakes, Judith S. Evans and Linda L. Brannon
The University of Texas at Austin
Austin, Texas 78712

Descriptors: Sentence Comprehension, Self-Embedding

Abstract

Sentences containing self-embedded (SE) relative clauses are generally believed to be difficult to understand because such clauses interrupt the clauses in which they are embedded. However, experimental results purporting to demonstrate this have confounded the SE or right-branching (RB) location of the relatives with their internal structure. Experiment I compared the comprehension difficulty of SE and RB object relative clauses on both phoneme monitoring and paraphrasing tasks; Experiment II made the same comparison for subject relatives. The results of both indicate that the internal structure of the relative clauses rather than interruption is the source of the difficulty.

Understanding Sentences with Relative Clauses[1]

David T. Hakes, Judith S. Evans and Linda L. Brannon
The University of Texas at Austin

Ever since the dawn of psycholinguistics, self-embedded (SE) sentences, like (1), taken from Chomsky and Miller (1963), have received an inordinate amount of attention. One reason for this attention was Chomsky's (e.g., 1956)

(1) Anyone who feels that if so-many more students whom we haven't actually admitted are sitting in on the course than ones we have that the room had to be changed, then probably auditors will have to be excluded, is likely to agree that the curriculum needs revision.

having called attention to such sentences in arguing that self-embedding was a fact about human languages that could not be captured by finite-state models. This results from the fact that such sentences contain constituents that are nested within other constituents such that portions of one constituent appear both to the left and to the right of another constituent. Self-embedding has, of course, the additional characteristic that the nested constituent is of the same structural type as that within which it is nested.

An additional reason for the attention SE sentences have received is that they are the canonical examples illustrating the need to distinguish between the linguistic character of sentences and the psycholinguistic character of the processing of these sentences. For while sentences with multiple self-embeddings are generally regarded as grammatical, it is clear that they are often unacceptable and are frequently very difficult to understand.

Considerable research in the past decade has been devoted to demonstrating this point and the related point that SE sentences are more difficult to understand than other sentences with comparable semantic complexity but different structure, i.e., that there is a special difficulty involved in understanding SE sentences. Miller and Isard (1964), for example, studied the memorization of sentences of constant length but varying numbers of SE relative clauses, finding that memorization became more difficult as the number of self-embeddings increased. Blumenthal (1966) found that subjects often paraphrase multiply SE sentences incorrectly; and Stolz (1967) found that only after considerable practice could subjects interpret such sentences correctly. Foss and Lynch (1969) and Hakes and Foss (1970) showed that on-line comprehension difficulty, measured by the phoneme monitoring task, was greater for doubly SE sentences than for semantically comparable right-branching (RB) sentences. And more recently, Blaubergs and Braine (1974) have reported that in a short-term memory probe task sentences with three or more levels of self-embedding were retained more poorly than comparable RB sentences.

Given the available evidence, there seems little doubt that multiply SE sentences are difficult to understand and more difficult than synonymous RB sentences, the sentence type with which they have generally been compared. It should be noted in passing, however, that not all the difficulty generally attributed to self-embedding need be attributed to that source per se. Stolz (1967), for example, found that SE sentences are considerably easier to understand if the predicate of each clause is appropriate to its own subject and inappropriate to the subjects of the sentence's other clauses. Further, it appears that many of the sentences on which the conclusion of self-embedding's difficulty has been experimentally based were simply bad sentences, independent of their being self-embedded. Our favorite example of this sort is (2), taken from the materials used by Foss and Lynch (1969). To us, this seems a blatant non-sequiter,

(2) The matron that the cat that the chauffeur kicked scared believed the gypsy.

there being no easily constructable reason why the chauffeur's kicking the cat has anything to do with the matron's believing the gypsy.[2] The sentence seems equally nonsensical in its "simpler" RB form, i.e., (3).

(3) The chauffeur kicked the cat that scared the matron that believed the gypsy.

But even if some of the difficulty generally attributed to self-embedding may better be attributed to other sources, the fact remains that such sentences are difficult to understand. The question is, how is this difficulty best accounted for?

The "standard" account of this difficulty is one proposed by Miller and his co-workers (see, e.g., Miller & Chomsky, 1963; Miller & Isard, 1964), based on the observation that SE sentences like (4) are more

(4) The prize that the ring that the jeweler that the man that she liked visited made won was given at the fair.

difficult to understand that RB sentences, like (5), that have the same

(5) She liked the man that visited the jeweler that made the ring that won the prize that was given at the fair.

number of relative clauses, and also based on formal considerations concerning how a sentence comprehension process might operate that makes only a single left-to-right pass through a sentence. Miller noted that each relative clause in a SE sentence interrupts the clause in which it is embedded, something not true of a RB structure. He proposed, essentially, that for a SE sentence the processing of the main clause must be suspended between the subject and the predicate in order to process the first relative clause. Processing of that clause must, in turn, be suspended in order to process the second relative clause, and so on for as many clauses as are embedded. Assuming a limited short-term processing memory that must retain parts of clauses until their completions are available, it is apparent that more than a few self-embeddings will overload the memory, leading to a loss of some of the clauses or, at least, of some of the information about which predicate goes with which subject.

Since Miller introduced the "interruption" hypothesis, it seems to have become almost a received truth; and alternative hypotheses about why SE sentences are difficult to understand have rarely been suggested. What we wish to do here is to propose such an alternative hypothesis.

As Miller noted, there is a difference in the location of SE and RB relative clauses. The former modifies the subject of the clause that dominates it, separating the subject from its predicate. The latter modifies the object of the dominating clause, occurring after the predicate of that clause.

However, for the SE and RB sentences that have entered into the experimental comparisons of these two sentence types there is an additional difference. These sentences have differed in the internal structure of the relative clauses used, a difference that has been completely confounded with the SE-RB difference. Essentially, the relative clauses used in SE sentences have been object relatives, ones in which it is the direct object of the clause that has been both fronted and relativized. The relative clauses used in RB sentences, however, have seen subject relatives, ones in which it is the subject of the clause that has been relativized. Thus, it is entirely possible that all the obtained results can be attributed to the difference in relative clause internal structure rather than to the difference between self-embedding and right-branching per se.

Bever (e.g., 1970, 1971) has suggested that any language has a customary or "canonical" order for constituents, one in which the superficial order coincides with the order in deep structure. For English, this appears to be the order of a "kernel" sentence or clause--Subject, Verb, and (optionally) Object. The results of a variety of experiments suggest that sentences are more difficult to comprehend if their constituents are not in canonical order than if they are. For example, innumerable studies have found that passive sentences, in which the constituents appear in the superficial order Object-Verb-Subject are more difficult to process than their active counterparts, in which the superficial order is the canonical order. To be sure, each comparison that is consistent with the generalization that violating canonical order leads to increased processing difficulty is also consistent with other interpretations. However, if the canonical order hypothesis is correct, then the difficulty of SE sentences lies not in the location of their relative clauses but, rather, in the internal structure of those clauses. The constituent order in a subject relative clause is, roughly, NP_1-rel-V-NP_2, where NP_1 (and the relative pronoun that is co-referential with it) is the subject of the clause, and NP_2 is its object. Thus, for such clauses constituent order is the canonical order. In an object relative clause, on the other hand, the fronting and relativizing of the clause's direct object yields an unusual order--NP_1-rel-NP_2-V, with both subject and object occurring before the verb, and the object before the subject. Thus, the canonical order hypothesis predicts that sentences containing object relative clauses will be more difficult than ones containing subject relatives, regardless of whether those clauses are self-embedded or right-branching.

Evaluating these two hypotheses--interruption and canonical order--requires breaking the confound that has been present in previous studies of self-embedding. This can be accomplished in either of two ways: (1) holding the internal structure of the relative clauses constant while varying their location; or (2) holding their location constant while varying their internal structure. We have chosen the first of these for the following reason.

Ideally, we wish to know not only whether there is a difference between sentences containing relative clauses as a function of either relative clause internal structure or of clause location (i.e., SE vs. RB); we would also like to know where in the course of comprehension processing the difficulty (and the difference in difficulty) appears. Post-sentence measures of comprehension difficulty, such as paraphrasing, probe latency or accuracy, etc., can yield information only about whether a difficulty difference exists. Evidence about where the difficulty arises appears to be obtainable only from an on-line measure, i.e., one that assesses difficulty during comprehension processing. For this reason we chose to use the phoneme monitor task (as well as a post-sentence measure of global difficulty, a paraphrasing task).

The phoneme monitor task requires that the subject, while trying to comprehend a sentence presented auditorily, listen in that sentence for a word beginning with a specified target phoneme and, when the target occurs, to respond to it as quickly as possible. Numerous experiments have shown reaction time in this task to be a sensitive measure of on-line processing load attributable to sentence structure (e.g., Foss & Lynch, 1969; Hakes, 1972; Hakes & Cairns, 1970; Hakes & Foss, 1970), to ambiguity (Foss, 1970; Foss & Jenkins, 1973; Swinney, 1974), and to a variety of other variables that affect comprehension difficulty.

On either the interruption or canonical order hypothesis, one would expect that a difference in difficulty would manifest itself within the relative clause itself. On the interruption hypothesis, it is the necessity of holding part of the dominating clause while processing the relative clause that causes the difficulty; hence, the effects of this holding of incompletely processed material should be evident during the processing of the relative clause. Similarly, on the canonical order hypothesis, the source of the difficulty is the necessity of dealing with an unusual constituent order, and that unusual order occurs within the relative clause. Thus, both hypotheses predict difficulty differences to be detectable during processing of the relative clause itself, either as a function of its location or as a function of its internal structure.

The problem that arises in attempting to evaluate these hypotheses by varying the internal structure of the relatives is just that any comparison of monitoring latencies within those relative clauses necessarily confounds clause structure with other variables. Consider, for example, sentences (6) and (7). Previous research (e.g., Foss, 1971) suggests

(6) The girl that the man liked moved to Chicago.

(7) The girl that liked the man moved to Chicago.

that monitoring latency is a complex function of where in a sentence (and, presumably, also within a clause) the monitor target occurs. For this reason, it would be desirable to hold the distance between the beginning of the clause (and/or the sentence) and the target constant. But if we decide to place the target on, say, the initial phoneme of the clause's fourth word, then the target in (6) is the /l/ in liked, while in (7) it is the /m/ in man. Since there is abundant evidence that monitoring latency is a function of the particular target phoneme (e.g., Savin & Bever, 1970) and also of the word containing the target (e.g., Foss, 1969; Cairns & Foss, 1971), any such comparison is confounded. Further, it appears that there is no way of eliminating such confounding. For this reason we have chosen to hold the internal structure of the relative clauses constant while varying their SE or RB location. Experiment I examines such SE vs. RB differences for object relative clauses, and Experiment II for subject relative clauses.

This solution is itself problematic in two ways. First, comparing sentences that all contain object relative clauses but differ in terms of whether those clauses modify the subject or the object of the dominating clause (e.g., Experiment I) necessarily involves comparing sentences that differ in meaning. The meaning of the relative clauses themselves can be constant across this comparison, and there is no particular reason to believe that the meaning difference between the sentences introduced by having relative clauses modify subjects or objects is a difference that systematically affects comprehension difficulty. But there is no way of avoiding this semantic difference without creating the structural confounding which we have hypothesized affects the interpretation of previous studies of self-embedding.

The second problem is of a somewhat different nature. Miller and Isard (1964) have suggested on the basis of their memorization results that there is a discontinuity in difficulty between one level of self-embedding and two or more levels. The evidence for such a discontinuity is weak and, further, is confounded, for as Miller and Isard decreased the number of SE object relative clauses their sentences contained they simultaneously increased the number of RB subject relatives. Nonetheless, there appears to be a pervasive intuition that only multiple self-embedding cause substantial comprehension difficulty.

For this reason, the strongest comparison would be between sentences with two or more object relatives in SE poisition and the same number in RB position, i.e., between two levels of self-embedding and zero levels.

Such a comparison is, however, impossible in principle, for a sentence with two object relative clauses attached to its direct object (i.e., RB) necessarily contains one level of self-embedding. Thus, the comparison has to be between one level of SE and none, or between two levels and one, etc. For this reason, we have chosen to examine sentences containing only single object relative clauses, located in either SE or RB position.[3]

In addition, on the possibility that differences in processing load might not become apparent until the overall load was relatively high, another variable--relative clause reduction--was included in both experiments. Fodor and Garrett (1967) have shown that deleting the relative pronouns from doubly SE sentences results in their being even more difficult to comprehend. Hakes and Cairns (1970) and Hakes and Foss (1970) have shown that this reduction also significantly increases RT in the phoneme monitor task. For this reason, each sentence in each experiment was presented with the relative clause either reduced or unreduced. This allows determining whether the reduction effect obtained previously with two levels of self-embedding occurs with only a single level.

Experiment I

Design and Materials. Twenty sentences were constructed, each of which contained a single object relative clause which could modify either the subject (SE) or object (RB) of the sentence's main clause, e.g., (8).

(8) (a) Before leaving for the theater, the Broadway star (that) the reporters had interviewed telephoned the producer to say he'd be late.

(b) Before leaving for the theater, the producer telephoned the Broadway star (that) the reporters had interviewed to say he'd be late.

The subject and object nouns of each sentence's main clause were interchangeable, so the relative clause always modified the same NP. This was done so that the object of the relative clause would be the same regardless of whether it modified the subject or object of the dominating clause. Since each sentence could be presented with the relative clause either reduced or unreduced and with the realtive clause modifying either the subject or object of the main clause, there were four versions of each sentence. The sentences were all complex in ways additional to their containing relative clauses, each comprising either three or four clauses.

Each sentence was constructed so that a word within the relative clause contained the target phoneme for monitoring. This was always the subject noun of the relative clause. Thus, in (8), the /r/ in reporters was the target phoneme. The target phoneme varied across sentences, but did not occur elsewhere in word-initial position in the sentence in which it was the target.

Twenty-eight filler sentences were also constructed, each of a degree of complexity comparable to that of the test sentences, but without any particular structural constraints observed in their construction. Of these, 15 contained monitoring targets, located at widely varying points and occurring in diverse types of words. The remaining 13 filler sentences contained no monitoring targets and were included to prevent the subjects' expecting such a target to occur in every sentence.

There was thus a total of 48 sentences, 20 of which contained object relative clauses. A single quasi-random order of the 48 sentences was constructed, and four tape recordings of the sentences in this order were made. The 20 test sentences were randomly divided into four sets of five sentences each. On any one tape, one subset occurred in each of the four possible combinations of relative clause location (SE vs. RB) and relative clause reduction (reduced vs. unreduced). The assignment of test sentence versions to tapes was accomplished by an ordered roll-over technique.

Tones of 1000Hz were recorded on the separate track of each tape, inaudible to the subjects, and conincident with the onset of each target phoneme. These were used to start the timer used to measure monitoring latency. Each sentence was preceded on each tape by an instruction specifying the target phoneme for that sentence. For the filler sentences that did not contain monitoring targets, the instruction was to listen for a phoneme that did not occur in word-initial position anywhere in that sentence. Instructions for the phoneme monitor and paraphrase tasks and four practice sentences were also recorded and dubbed onto the beginning of each tape.

To summarize, the design included four groups, defined by the four tapes, i.e., by the assignment of test sentences to conditions. Within subjects, there were two values of each of two variables: relative clause location (SE vs. RB) and relative clause reduction (reduced vs. unreduced). The design was thus a 4 x 2 x 2 mixed factorial.

Procedure. Subjects were assigned to groups randomly and were all tested individually by the same experimenter. The instructions and sentences were presented binaurally through headphones at a comfortable listening level. The general instructions indicated that the subject was to try to comprehend each sentence and was to: (a) listen for a word in each sentence beginning with the sound specified by the instruction preceding that sentence; (b) respond by pressing a button as quickly as possible if such a word were heard; and (c) say in his own words, after the sentence was finished, what the sentence meant. The instructions explicitly discouraged verbatim repetition of the sentences, and, if necessary, subjects were further instructed on this point after the practice sentences. The four practice sentences were then presented, followed by the 48 test and filler sentences. The button used for the monitoring response was mounted on a panel angled toward the subject at 30° above the horizontal plane. Subjects responded with whichever digit they preferred. Paraphrase responses were made orally and tape recorded for later analysis.

Subjects. Forty students from the freshman-level introductory psychology course at The University of Texas at Austin served as subjects. Participation in experiments was a course requirement, but the subjects were not required to serve in this particular experiment. All were native speakers of English, and none had previously served in any experiment involving sentence comprehension, phoneme monitoring or paraphrasing. Approximately equal numbers of males and females served, though subjects were neither selected nor assigned to groups on this basis.

Results and Discussion

Phoneme Monitoring. Individual monitoring latencies were converted to speed scores by reciprocal transformation. For each subject, four mean speed scores were computed, one for the sentences of each of the four conditions. The means of these mean speed scores, taken across groups, are presented in Table 1 and indicate that monitoring was faster when the clauses occurred in SE location than when they occurred in RB location. Further, as anticipated, monitoring was faster when the relative clauses were unreduced than when they were reduced.

Insert Table 1 about here

The mean speed scores were subjected to a Groups x SE vs. RB x Reduced vs. Unreduced analysis of variance. The main effect for SE vs. RB was significant, $F(1,36) = 5.499$, $p < .05$, as was the main effect for

Reduced vs. Unreduced, $F(1,36) = 24.617$, $p < .001$. The main effect for Groups was far from significant, $F < 1$, and interactions approached significance. The results of the analysis thus confirm the effects discovered by inspection.

The effect attributable to relative clause reduction indicates that deleting the relative pronouns from object relative clauses substantially increases on-line processing difficulty even when the sentences upon which this variation is worked are less complex than the double self-embeddings used in earlier experiments.

The effect attributable to relative clause reduction indicates that deleting the relative pronouns from object relative clauses substantially increases on-line processing difficulty even when the sentences upon which this variation is worked are less complex than the double self-embeddings used in earlier experiments.

The effect of main interest is, of course, that attributable to the SE or RB location of the relative clauses, for there the interruption hypothesis clearly predicts that monitoring should be slower for SE clauses than for RB clauses. The canonical order hypothesis, on the other hand, makes no clear prediction for this comparison. If everything else were equal, it should predict no difference. But everything else is seldom equal, and here we might expect the processing load to vary as a function of the amount of other processing (other, that is, than of the relative clause itself) occurring at the point where the target occurred. Since RB relative clauses always occurred somewhat later in their respective sentences than SE relative clauses, the total processing load might very well vary with this difference. But without a detailed theory of such other processing it would be difficult to predict a difference in either direction.

The results indicate, of course, that processing load was greater for RB clauses than for SE clauses. Since the interruption hypothesis unambiguously predicts an effect in the opposite direction, it is clear that these results lend no aid or comfort to that hypothesis.

Paraphrasing. Each test sentence was divided into segments corresponding roughly to surface structure clauses, there being either three or four such segments per sentence. The paraphrase responses were scored on the basis of the number of such segments accurately represented in the response. To be scored correct, both the internal grammatical relations of a segment and also its relations to the sentence's other segments had to be represented accurately. Lexical substitution was liberally allowed, except for the substitution of an indefinite term for a definite (e.g., someone for the Broadway star). Verbatim repetitions of the sentence were not credited.

Each paraphrase response was transcribed and was scored independently by two judges. Each judge scored all paraphrases of all versions of a given test sentence before proceeding to the next sentence. Differences in the scores assigned by the two judges were resolved by discussion (sometimes acrimonious) after all scoring was completed.

Since the number of segments scored varied across sentences, a subject's accuracy score for a given sentence was taken as the proportion of that sentence's segments accurately paraphrased. Four mean accuracy scores were then computed for each subject by taking the mean proportion correct across the sentences in each of the four conditions. The means of these mean accuracy scores, taken across groups, are presented in Table 1.

The mean accuracy scores were subjected to analysis of variance. The main effect for SE vs. RB location was highly significant, $\underline{F}(1,36) = 7.653$, $\underline{p} < .01$, paraphrasing being more accurate for SE object relative clauses than for the same clauses in RB location. This result is consistent with that obtained for the phoneme monitor task, but inconsistent with the interruption hypothesis.

The main effect for relative clause reduction was also highly significant, $\underline{F}(1,36) = 10.067$, $\underline{p} < .01$. But, paradoxically, the effect was in the wrong direction--paraphrasing was better for sentences with reduced relative clauses than for ones where the relative clauses were unreduced. This result is consistent neither with the results for the phoneme monitor task nor with previous paraphrasing results. The interaction of relative clause reduction with SE vs. RB location was not significant, $\underline{F} < 1$. The only other effect in the analysis that did reach significance was the Groups x Reduced vs. Unreduced interaction, $\underline{F}(3,36) = 7.630$, $\underline{p} < .01$, indicating that for one assignment of test sentences to conditions paraphrasing was more accurate when the relative clauses were unreduced than when they were reduced.

In summary, then, the results for both tasks are consistent in indicating that comprehension difficulty is less for SE object relative clauses than for RB object relative clauses. This is not inconsistent with the canonical order hypothesis, though not predicted by it. It is, however, inconsistent with the interruption hypothesis. The conventional wisdom that SE sentences are more difficult to understand than almost anything else is challenged.

The effects for relative clause reduction are more complex. The phoneme monitoring results are consistent with those of previous research in finding that reduction increases comprehension difficulty. The paradoxical result is that paraphrasing accuracy was greater for the sentences with reduced relatives. Fortunately, further data are available which shed some light on this result.

A second experiment has been performed using the same test sentences as Experiment I. The major difference was that an on-line probe task was substituted for the phoneme monitor task. During sentence presentation, at the same point at which the monitor target occurred in Experiment I, a word appeared on a screen in front of the subject. The subject's task was to decide as quickly as possible whether that word had occurred in the sentence. We will not detail the results for that task here, noting only that the same trends appeared as did for the phoneme monitoring task. These trends were, however, far from significant, owing largely to inordinately large within-condition variance.

The results of interest in the present context are those for the paraphrase task. These data were dealt with in a manner identical to that described earlier for Experiment I, except that one of the two judges was different. The results, in terms of mean proportion accuracy scores, are presented in Table 2. As in Experiment I, the effect for SE vs. RB location was highly significant, $\underline{F}(1,36) = 15.010$, $\underline{p} < .001$, again indicating that paraphrasing accuracy was greater for SE relative clauses than for RB relative clauses. The effect for relative clause reduction did not even approach significance, $\underline{F}(1,36) = 1.004$, $\underline{p} = .34$.

Inspection revelas that this insignificant difference was in the direction found in earlier experiments with SE sentences, not in the direction found in the paraphrasing data of Experiment I.

Insert Table 2 about here

It appears, then, that the unexpected effect found for relative clause reduction in the paraphrasing data of Experiment I is best attributed to chance; it is not a replicable effect. It is, however, unclear why the effect found in earlier experiments--increased difficulty with reduction--was not obtained here. Our best guess is that the paraphrase task is not sufficiently sensitive to small differences in comprehension difficulty to consistently reveal such an effect. (See Hakes, 1972, for a discussion of some of the relevant shortcomings of this task.) Such insensitivity may well have been compounded by some characteristics of the particular sentences used here. But if so, it is not clear what those might be. In any event, it appears that the aberrant paraphrasing effect found in Experiment I is not a major cause for concern.

Experiment II

Experiment I has demonstrated that, for object relative clauses, interruption of the dominating clause does not result in increased comprehension difficulty, thus disconfirming the interruption hypothesis. But the interruption hypothesis does not apply just to object relative clauses. If the hypothesis were true, interruption of a dominating clause should result in increased comprehension difficulty, regardless of the internal structure of the embedded clause. Thus the hypothesis makes the same prediction concerning SE or RB location for subject relative clauses as it does for object relative clauses.

The canonical order hypothesis, on the other hand, predicts that if the SE or RB location of the clauses remains constant, object relative clauses should cause greater comprehension difficulty than subject relatives. We have earlier noted one problem making difficult a direct test of this prediction--that the differing internal structure of subject and object relative clauses precludes a non-confounded test with on-line measures of comprehension difficulty. There is, in addition, a further problem. Since the canonical order hypothesis makes a strong prediction about the effects of relative clause internal structure only when all relative clauses are either SE or RB, confounding occurs for any direct test, even those made with post-sentence comprehension tasks. Varying the internal structure of relative clauses necessarily involves varying their semantic content as well as the content of the sentences that contain them. Admittedly, as mentioned earlier, the comparison performed in Experiment I was confounded in a similar way, for even though the content of the relative clauses remained constant across SE and RB locations, the content of the sentences containing them varied somewhat systematically. Still, it seems intuitively far worse to systematically vary the content of the relative clauses themselves; and this consideration makes any direct test of the canonical order hypothesis risky. It seems that the better strategy is to provide direct tests of the interruption hypothesis and indirect ones of the canonical order hypothesis.

Experiment II provides a further case of this sort, comparing sentences containing subject relative clauses in either SE or RB location.

Method

Design and Materials. The materials here paralleled as nearly as possible those of Experiment I, except for the internal structure of the relative clauses used. Again, 20 test sentences were constructed, each of which contained a single subject relative clause which could modify either the subject or the direct object of the sentence's main clause. Whenever possible, the sentences of Experiment I were used, modified in this way.

The verb in the relative clause was always in the past progressive tense, allowing us to present the sentences with the relative clauses either reduced or unreduced. There is a general rule of English syntax which prohibits reducing a subject relative clause if its main verb is nonprogressive. One reason for this appears to be that when such a relative is reduced the fact of its existence is obscured, making miscomprehension of the sentence likely (see Bever & Langendoen, 1971).

The filler sentences were identical to those of Experiment I, as were all details of tape construction and the placement of tones for phoneme monitoring. The monitoring target for each test sentence was the initial phoneme of the head noun of the relative clause's direct object. Here, as in Experiment I, there were four groups of subjects, defined by the assignment of test sentences to the SE and RB and reduced and unreduced conditions. Within subjects, these four conditions were combined orthogonally.

Procedure. The procedure was identical in all respects to that of Experiment I.

Subjects. Forty subjects, having the same general characteristics as those used in Experiment I, participated. Ten were assigned to each group according to a quasi-random schedule.

Results and Discussion.

Phoneme Monitoring. Reduction of the monitoring data to mean speed scores was accomplished in the same manner as in Experiment I, with one exception. After data collection was completed it was discovered that one test sentence had inadvertently been recorded incorrectly on one tape. Data for this sentence were eliminated for all groups, with the result that the mean speed scores for the conditions into which that sentence entered were based on one fewer latency than the other conditions. The means of these mean speed scores, taken across groups, are presented in Table 3.

Insert Table 3 about here

Overall, responding here was somewhat faster than in Experiment I. This suggests that the subject relative clauses caused less comprehension difficulty than the object relatives, though it might also reflect other differences between the sentences of the two experiments, e.g., the difference in monitor target location. The difference between SE and RB conditions is in the same direction as in Experiment I, monitoring being slightly faster when the relative clauses were self-embedded. The effect is, however, far smaller than that obtained with object relative clauses. Monitoring was also slightly faster when the relative clauses were reduced than when they were unreduced, but again the effect is miniscule.

Analysis of variance of the mean speed scores bears out the impressions gained by inspection. The main effects for SE vs. RB location and

Reduced vs. Unreduced were both far from significant, $F < 1$ in both cases. The only effect to reach significance was that attributable to the Groups x SE vs RB interaction, $F(3,36) = 3.880$, $p < .02$, and in the absence of a main effect for SE vs. RB location this is of little consequence.

It appears, then, that no differential on-line processing load resulted from differences in either the SE or RB location of the subject relative clauses or from whether they occurred in complete form or with the relative pronouns and auxiliary verbs deleted. The latter finding suggests that the reduction effects found for object relative clauses are not entirely general, i.e., deletion of an optional cue to sentence structure does not *always* result in increased on-line processing load. The fact that processing load was not greater when the relative clauses were SE is inconsistent with the interruption hypothesis. In addition, informal comparison of these results with those of Experiment I suggests that subject relative clauses caused less on-line comprehension difficulty than object relative clauses, regardless of whether they were SE or RB. This, of course, is the result predicted by the canonical order hypothesis.

Paraphrasing. The paraphrasing data were treated in a manner identical to that described for Experiment I, with the same exception noted for the phoneme monitoring data for Experiment II. Again, two judges independently scored the paraphrase data, using the same procedures and criteria as in Experiment I.

The paraphrasing data are presented in Table 3 in terms of mean proportion accuracy scores, taken across groups, for the four conditions. The overall level of accuracy is considerably higher than that found in Experiment I for sentences containing object relative clauses, again suggesting that subject relatives cause less comprehension difficulty than object relatives. Accuracy appears to have been somewhat higher for subject relatives in SE location than for the same relative in RB location, and also somewhat greater when the relatives were unreduced than when they were reduced.

Analysis of variance of the paraphrasing data revealed that the main effect for SE vs. RB location was significant, $\underline{F}(1,36) = 4.503$, $\underline{p} < .05$. The main effect for Reduction was not significant, $\underline{F} < 1$. Somewhat complicating the interpretation of these effects is the fact that the main effect for Groups was significant, $\underline{F}(3,36) = 3.133$, $\underline{p} < .05$, as were the interactions of Groups x SE vs. RB location, $\underline{F}(3,36) = 9.463$, $\underline{p} < .001$, and Groups x SE vs RB x Reduction, $\underline{F}(3,36) = 3.126$, $\underline{p} < .05$. Inspection of the data suggests that these interactions are primarily due to unusually low accuracy (.42) for one group for the sentences with reduced relative clauses in SE location. Performance for the set of sentences under other conditions for other groups of subjects was not atypical, suggesting that the cause of the atypical performance in that one cell was not the sentences themselves. However, the subjects in the group that contributed the atypical performance tended to be less accurate for all conditions than subjects in the other three groups. This suggests that it was the subjects that were abberant. Inspection of the phoneme monitoring data failed to yield any evidence of unusual performance for that group.

The paraphrasing data are, then, consistent with the monitoring data in showing no consistent effect for relative clause reduction. In addition, they reveal a significant difference between SE and RB relative clauses that was not apparent in the monitoring data. And, like the monitoring data, the overall level of paraphrasing accuracy suggests that the subject relative clauses of Experiment II caused less comprehension difficulty than the object relative clauses of Experiment I.

The fact that SE relatives yielded better performance than RB relatives for the paraphrase task but not for the phoneme monitor task is somewhat puzzling, for in general paraphrasing is a considerably less sensitive test than monitoring. While the paraphrasing effect is what might be expected on the basis of Experiment I's results, it is at least mildly surprising that the effect should appear for the less sensitive, but not for the more sensitive task.

It is possible that the source of the discrepancy lies in the phoneme monitor task as it was used in this experiment. The monitor targets in the subject relative clauses occurred later than was the case for the object relative clauses in Experiment I (i.e., further from the beginning of the clause). It is possible that the difference in processing load between SE and RB locations is relatively transient and had disappeared by the point of the monitor target. But if there is an effect and it is that transitory, then it is a bit difficult to see how it could have a significant effect on an insensitive global comprehension difficulty measure like the paraphrase task.

Alternatively, the source may lie in the paraphrase task. It is by no means clear that the scoring criteria for paraphrase accuracy are ideally suited to the task of assessing comprehension difficulty for the kinds of sentences used here. In addition, the general insensitivity of the paraphrase task to differences in comprehension difficulty is undoubtedly partly attributable to (relatively) low reliability. But low reliability does not insure that a true difference will be underestimated. Rather, it insures a certain capriciousness in the relationship between the true and observed differences, with the result that the true difference is sometimes overestimated. The effect observed there may be such an instance.

General Discussion

The relative clause reduction variable was included in the present experiments mainly to increase the overall comprehension difficulty of the sentences. We had anticipated that difficulty differences between SE and RB sentences might not appear until overall difficulty was high. Clearly, this assumption was disconfirmed, for in neither experiment was there an interaction between reduction and relative clause location. Whatever differences in comprehension resulted from these two manipulations were quite independent. Difficulties attributable to different sources need not be cumulative.

The fact that relative clause reduction yielded a significant increase in comprehension difficulty for the object relative clauses in Experiment I suggests that the reduction effects found in previous experiments using doubly self-embedded sentences are not attributable to the fact that two levels of self-embedding were involved. The effect appears to be more general than that.

The mixed results of Experiment II suggest, however, that there may be some limits on the generality of the effects of deleting optional cues on comprehension difficulty. Given the mixed character of these results, it would be unwise to place much reliance on them as either supporting or denying generality to the effect. It may well be, however, that reducing a subject relative clause by deleting the relative pronoun and auxiliary verb has effects different from those of reducing object relative clauses or NP-object complement constructions (Hakes, 1972). There may be changes in the intonation and stress pattern of reduced subject relative clauses that provide the same information as the relative pronoun and auxiliary verb. Or it may be that hearers can make use of the -ing ending on the relative clause's verb to make the same inference about the structure of the input string. But whether an account in terms of alternative cues to subject relative clause structure is necessary will depend on further data establishing more clearly whether or not the comprehension of sentences with subject relatives is deleteriously affected by relative clause reduction.

The results of main concern are, of course, those pertaining to the differences between self-embedded and right-branching relative clauses. We began with the question, why are self-embedded sentences difficult to understand? The answer seems to be that the question involves a mistaken presupposition, for our results make clear that SE sentences are not necessarily more difficult to understand than other sentences containing relative clauses. In fact, the results of both experiments suggest that, if anything, sentences with SE relative clauses are less difficult than sentences in which the same relative clauses are RB. Thus, the better question is, why have previous experiments concluded that self-embedding caused difficulty? And here the answer seems to be that that conclusion rests on the comparison of sentences containing SE object relative clauses with sentences containing RB subject relatives, confounding the

SE vs. RB difference with the difference in the internal structure of the relative clauses.

Given this confounding, however, these studies have found that sentences with SE object relatives are more difficult than sentences with RB subject relatives. Since the present experiments have shown that the SE vs. RB difference itself cannot account for the difficulty difference, it follows logically that the difficulty must have something to do with the difference between object relative clauses and subject relative clauses.

That conclusion would be more palatable if, in addition to an indirect logical argument, we could present a direct empirical argument. Unfortunately, for the reasons indicated earlier, such a direct argument cannot be made, for the relevant comparisons are necessarily confounded. Comparison of the results of Experiments I and II is suggestive, for it does appear that for both the phoneme monitor and paraphrase task subject relative clauses are less disruptive of comprehension processing than object relative clauses. Still, we are reluctant to place great confidence in that comparison, both because of the kinds of confounding discussed earlier and because the comparison involves sets of data collected from different subjects at different times by different experimenters. We have, for these reasons, refrained from performing any statistical analysis of the between-experiment comparison. However unsatisfactory this state of affairs is, we feel we must rest the case with a strong argument against the difficulty being attributed to the fact of self-embedding per se, with its consequent interruption of a dominating clause, and a weaker argument in favor of its being attributed to the internal structure of object relative clauses.

We suggested earlier that the reason object relative clauses resulted in increased comprehension difficulty was that the order in which their constituents occurred was an unusual order. To the extent that the dominant comprehension processing strategies depend on canonical constituent ordering, other orders will result in increased difficulty. No doubt there are other accounts that could be suggested for the differential difficulty of object and subject relative clauses; but the canonical order hypothesis seems a reasonable candidate. Whatever the correct account may be, one thing is clear: any theory of comprehension processing that accords a major role in creating comprehension difficulty to the interruption of processing of one constituent for the processing of an embedded constituent must surely be false.

References

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.) *Cognition and the development of language*. New York: Wiley, 1970. Pp. 279-362.

Bever, T. G. The nature of cerebral dominance in the speech behaviour of the child and adult. In R. Huxley & E. Ingram (Eds.) *Language acquisition: Models and methods*. New York: Academic Press, 1971. Pp. 231-255.

Bever, T. G., & Langendoen, D. T. A dynamic model of the evolution of language. *Linguistic Inquiry*, 1971, 2, 433-464.

Blaubergs, M. S., & Braine, M. D. S. Short-term memory limitations on decoding self-embedded sentences. *Journal of Experimental Psychology*, 1974, 102, 745-748.

Blumenthal, A. L. Observations with self-embedded sentences. *Psychonomic Science*, 1966, 6, 453-454.

Cairns, H. S., & Foss, D. J. Falsification of the hypothesis that word frequency is a unified variable in sentence processing. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 41-43.

Chomsky, N. Three models for the description of language. *I.R.E. Transactions on Information Theory*, 1956, 2, 113-124.

Chomsky, N., & Miller, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.) *Handbook of mathematical psychology*, *Vol. II*. New York: Wiley, 1963. Pp. 269-322.

Fodor, J. A., & Garrett, M. Some syntactic determinants of sentential complexity. *Perception and Psychophysics*, 1967, 2, 289-296.

Foss, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 457-462.

Foss, D. J. Some effects of ambiguity upon sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 699-706.

Foss, D. J. On the time-course of sentence comprehension. Paper presented at the C.N.R.S. Colloquium, "Problemes actuels en Psycholinguistique," Paris, 1971.

Foss, D. J., & Jenkins, C. M. Some effects of context on the comprehension of ambiguous sentences. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 577-589.

Foss, D. J., & Lynch, R. H., Jr. Decision processes during sentence comprehension: Effects of surface structure on decision times. Perception and Psychophysics, 1969, 5, 145-148.

Hakes, D. T. Effects of reducing complement constructions on sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 278-286.

Hakes, D. T., & Cairns, H. S. Sentence comprehension and relative pronouns. Perception and Psychophysics, 1970, 8, 5-8.

Hakes, D. T., & Foss, D. J. Decision processes during sentence comprehension: Effects of surface structure reconsidered. Perception and Psychophysics, 1970, 8, 413-416.

Miller, G. A., & Chomsky, N. Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.) Handbook of mathematical psychology, Vol. II. New York: Wiley, 1963. Pp. 419-492.

Miller, G. A., & Isard, S. Free recall of self embedded English sentences. Information and Control, 1964, 7, 292-303.

Savin, H. B., & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 295-302.

Stolz, W. S. A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 867-873.

Footnotes

[1] This research was supported by Grant OEG-0-72-3941 from the National Institute of Education. The authors wish to thank Donald J. Foss for his comments.

[2] Foss (personal communication) has furnished us with the following "context" for the sentence, which we quote without comment:

> Once upon a time there was a rich old woman who was the target of many people out to get her money. The lady had become quite a skeptic about human nature as a result of this, and she was always on the lookout for people out to take her money. One day some casual "friends" (who were really part of a ring desiring to bilk her out of her funds) talked her into going to a seance. The plan was to get the old lady to trust the medium (who dressed like a gypsy) and then to involve her in a scheme that would surely cause her to lose her money. On the first visit to the gypsy the old woman was her skeptical self. However, among the predictions that the gypsy made was the following: that the old woman would, very soon, be frightened by someone that she loved. Now the old woman loved no one but her cat, and so she know that this prediction could not come true. However, on leaving the seance and entering her car, she was met with a horrible sight. Her chauffeur was mistreating her cat to such an extent that the cat got angry and turned on the old woman; she was indeed frightened by the experience. The matron that the cat that the chauffeur kicked scared believed the gypsy. And, of course, she later lost everything. Moral: One person's non-sequiter can be another person's punch line.

[3] Note that even if Miller and Isard's claim about a discontinuity is correct it is a claim about the amount of processor overload that is sufficient to cause a complete breakdown of comprehension processing. Since the phoneme monitor task appears to be sensitive to differences in processing load at levels lower than those resulting in such a breakdown, there may be no problem at all.

Table 1

Phoneme Monitoring Speed and Paraphrasing

Accuracy: Experiment I

Condition	Monitoring Speed	Paraphrasing Accuracy
Self-Embedded		
Unreduced	1.997	.451
Reduced	1.811	.514
Right-Branching		
Unreduced	1.934	.409
Reduced	1.720	.460

Table 2

Paraphrasing Accuracy: Supplementary Experiment

	Unreduced	Reduced
Self-Embedded	.570	.560
Right-Branching	.508	.489

Table 3

Phoneme Monitoring Speed and Paraphrasing Accuracy: Experiment II

Condition	Monitoring Speed	Paraphrasing Accuracy
Self-Embedded		
Unreduced	1.989	.679
Reduced	2.021	.646
Right-Branching		
Unreduced	2.009	.625
Reduced	1.984	.628

APPENDIX III

Effects of Context upon the Processing of Lexical Ambiguities during Sentence Comprehension

David A. Swinney

EFFECTS OF CONTEXT UPON THE PROCESSING
OF LEXICAL AMBIGUITIES DURING
SENTENCE COMPREHENSION

by

DAVID ALLEN SWINNEY, A.B., M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August, 1974

ACKNOWLEDGMENTS

There are a large number of acknowledgments which must be made concerning the work associated with this dissertation. This type of work is rarely accomplished without great intellectual debts to others, and this paper is certainly no exception; I am both grateful to those who have helped me, and very pleased to acknowledge these debts.

First, and most importantly, this work would not and could not have been accomplished without the patient, selflessly given, help and guidance of Dr. David T. Hakes. There is absolutely no way to adequately express the enormity of my gratitude and my debt to him.

In addition, I would like to express my sincere thanks to Dr. Donald Foss for his guidance and interest in this work from its outset, and for his interested and professional counsel in all related matters.

Finally, I wish to express warm thanks and appreciation to my good friends and teachers who have encouraged and guided me throughout this work.

This research has been supported by a predoctoral research fellowship from the National Institute of Mental Health and a grant (OEG-0-72-3941) from the National Institute of Education to David T. Hakes. Sincere thanks are given to Dr. Hakes, the National Institute of Education, and the National Institute of Mental Health.

D.A.S. July, 1974

TABLE OF CONTENTS

Page

Acknowledgments iii

List of Tables, Figures and Appendices v

Chapter I - INTRODUCTION

- Introduction 1
- Overview of the Language Processing Sequence Occurring During Language Comprehension 3
- General Statement of the Experimental Problem 7
- Rationale for the Study of Ambiguity 10
- Experimental Studies of Ambiguity
  - a. Effects of unbiased ambiguity on sentence processing 12
  - b. Effects of post-disambiguating context upon processing of ambiguities in sentences 16
- Studies of Sentential Processing of Ambiguities Involving Prior Context 18
- Discussion and Evaluation of Methodology and Materials Used in Studies of Prior Context and Ambiguities 24

Chapter II - METHOD

- Design and Materials 32
- Subjects 36
- Procedure 37

Chapter III - RESULTS 39

Chapter IV - DISCUSSION

- I 49
- II 57

Chapter V - SUMMARY, CONCLUSIONS, FUTURE RESEARCH 72

Appendices 78

Bibliography 94

Vita 98

LIST OF TABLES, FIGURES AND APPENDICES

Page

TABLES

1. Mean reaction times, in msec, to the six conditions of the ambiguity by context interaction. 40

2. Mean reaction times (msec) to the six ambiguity by context conditions for (a) subjects who noticed the presence of ambiguities in the materials and (b) subjects who did not notice ambiguities in the materials. 65

FIGURES

1. Mean reaction times (msec) to monitor in sentences containing Ambiguous and control (Nonambiguous) words as a function of Context, for subjects who were either Aware or Unaware of the presence of ambiguities in the materials. 47

APPENDICES

A. Frequency (Thordike-Lorge Count) and length (in syllables) of ambiguous words and unambiguous control words used in experimental sentences (TI). 78

B. List of experimental sentences in schematized form. 80

C. Filler sentences with monitor target occurring early in the second sentence. 83

D. Filler sentences without target present. 85

E. Sample response sheet. 87

F. Instructions given to Ss. 88

G. Source table for ambiguity x control x groups analysis of variance. 91

H. Summary of results of Neuman-Keuls Test performed on the ambiguity x context interaction 92

I. Distance (in syllables) between context and text item. 93

CHAPTER I

INTRODUCTION

The effect of differing types of 'context' upon human cognitive processing has been a focus of interest throughout the history of the study of cognitive processes in psychology. The range of variables which might be called 'contexts' (as well as the cognitive processes studied) in this respect is enormous. It includes, among others, such variables as 'type of field' in figure-ground perceptual studies (Woodworth & Schlosberg, 1954) and 'preparatory set' and 'expectancy' in studies of association, problem solving (e.g., Maier, 1930), signal detection (e.g., Miller, Heise, & Lichten, 1951), and memory (e.g., Norman & Rumelhart, 1970). Such concern about the interaction of contextual information with information more commonly considered to be directly involved in the various cognitive processes is understandable, for man rarely uses these processes in isolation. Given the general findings of the kinds of studies cited above, a claim that 'contexts' can have an effect upon the outcome of cognitive processing seems indisputable. What is less certain, however, is exactly how 'contexts' interact with these processes (i.e., the nature of the interaction as opposed to its outcome).

The study of both information interaction and expectancy in language processing, then, is just one facet of a more general concern in cognitive psychology. A focus on language, however, is the major concern of this paper. While the study of such effects on language comprehension and production is certainly of importance in its own right, the vehicle to be utilized for such study in this paper is one which is also of general interest in other areas of cognitive psychology.[1] Thus, hopefully, this study of language processing will indicate useful directions for the study of cognitive processes in general, as well as provide some answers to questions concerning language processing per se.

Language comprehension is a process which requires the interactive processing of various types of linguistic information comprising an utterance.[2] In such processing, information from all levels of linguistic analysis (i.e., semantic, syntactic, phonological, etc.) is involved. One way in which to consider such information interaction is to regard the processing of any one piece of information as providing a *context* for operations on another. There are obviously many types of such potential contexts involved in language processing: the surface and deep structures of an utterance; the physical state of the environment when the utterance is heard; the 'theme' of a discussion; prior knowledge and expectancies about the world; specific sounds, words, and phrases preceding or following the utterance; and so on. Further, while it may appear reasonable to assume that 'context' will have some effect upon almost every 'level' of processing in any model of language comprehension, there is little empirical support for such a generalization;

---

[1] This 'vehicle' is ambiguity, a topic to be discussed at length, below.

[2] In the paradigm case, linguists have shown syntax to interact with word meaning in the derivation of meaning for a sentence; old associationistic notions that sentence meaning derives from a simple summation or intersection of each word meaning have long ago been shown to be false. (Hence, *house cat* does not mean the same as *cat house*, and *the alleged dictator* differs from *the dictator alleged*.)

the loci of context effects are virtually unknown. The problem of how and where semantic information, such as that assumed to exist for individual lexical items, interacts with similar information about other lexical items during language processing is a part of the context/information interaction problem which must be solved in order to even begin an accurate description of the language processing mechanism. This paper explores a portion of that problem.

Overview of the Language Processing Sequence Occurring during Language Comprehension

The study of context effects during language comprehension necessarily involves certain assumptions concerning the approximate sequence of language processing performed by a listener. The assumptions of interest at this point are general in nature and are more concerned with suggesting the relative time-course of processing than with establishing any firm model of language comprehension.

The first assumption is that the language processor, upon hearing an input string of speech sounds, segments this string into units roughly corresponding to words. These units are then 'accessed' or 'looked up' in an internal lexicon (memory) to find information related to them. Clearly, such an assumption raises more problems than it solves. For instance, there is the problem of how speech is perceived and how segmentation into words occurs. Liberman and his associates (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) have detailed many of the problems yet to be explained in this area. The task is apparently accomplished in some fashion, however, and with that in mind, any attempt at explanation of how the process is accomplished is elided here.

It is further assumed that information about a 'word' is somehow related to its phonological code and that the act of perceiving the physical speech sounds and segmenting them into a unit corresponding to a word is the act of obtaining access to the information associated with that word. Unfortunately, again, the nature of the process by which such access occurs must be left unexplained; there are too many possibilities and too little data to decide among them. For example, the recent work of Meyer and his associates (e.g., Meyer & Schvaneveldt, 1971) utilizes a notion of 'activation' to explain access of related information. This is a concept which apparently derives from the ideas of James Mill and other British associationist philosophers and which has been carried over into more current views of memory function (e.g., Hebb, 1949). In its simplest form it is an essentially passive process, whereby everything linked with (or 'associated to') the phonological code simply achieves heightened (neuronal?) excitation and, hence, becomes more available to whatever processes are going on. On the other hand, Neisser (1966) suggests that retrieval of information associated with an item in long-term memory is essentially a process of active reconstruction of that item--where the method by which the reconstruction is attempted determines exactly what information is made available. There are other possible models, but, as mentioned, there are at present few data on which to base a decision among them. Further, while knowledge concerning such models is intrinsically interesting and important, it is not essential to the major pursuits of this paper.

As a final problem worthy of brief mention in relation to these initial assumptions, the question of exactly what information is stored 'with' (metaphorically speaking) the lexical entry and what information is actually retrieved when the word is accessed requires comment. It is assumed that the information stored with the phonological code for a word comprises the meaning of the word in its most general sense: it includes, perhaps, syntactic features (e.g., class occurrence restrictions) and semantic features (conceptual primes which 'define' the word).[3] However, this general claim is clearly an empty one insofar as it fails to describe what of the possible related information is retrieved. The problem is not trivial. All words (with the possible exception of grammatical formatives) have a potentially large amount of information associated with them. The problem comes in explaining and describing what part of this information is made available in different situations. As a simple example of the problem, the word 'doctor' has been shown to yield 'nurse' as a high associate, and hence can be assumed to be highly related to that term and the semantic information which comprises it. However, the word 'doctor' would be unlikely to yield 'over four feet tall' as a high associate, even though this information can be clearly related to doctor.' In fact, to further compound the problem, all words would seem to be related to each other, if by no other means than the relationship: 'the set of all words.' (Of course, the problem is not one of 'words' being related to each other, but, rather, of the relationships of the information--assumed to be in some form of conceptual primes--which comprise the meaning of these words.) As all information is apparently related to all other information, an unconstrained, passive 'activation' system must fail. In such a system, with no constraints on 'spreading activation' all information would become activated upon access of one word and thus the notion of activation could not account for how the meaning of a single word becomes known. In order to avoid this pitfall, researchers such as Meyer (1971) have utilized a concept of 'semantic distance' in an attempt to constrain and quantify the extent to which activation occurs. By this account, semantically 'closer' items become more activated than more 'distant' items. However, this does not solve the problem presented earlier, but merely acknowledges it through the use of labels; 'semantic distance' has, at best, been an ill-defined concept. In short, this brings us back to the point of the original question concerning what information is made available when the phonological code is accessed. Clearly it is a problem worthy of study, and it is hoped that the present research may provide some answers. However, as with the other questions raised by the basic assumptions currently under discussion, this problem will necessarily remain unanswered for the present; constraints on the possibilities for models of storage and access are virtually non-existent at this point. Before leaving the topic completely, however, a couple of additional points need to be made. First, some set of semantic information appears to be regularly made available when the phonological code for the word is encountered in language comprehension. Second, as will be discussed below, more information appears to be made available upon occurrence of certain types of ambiguous words than for unambiguous ones.

---

[3] This is the case for lexical morphs, or lexical formatives. Grammatical formatives may not carry semantic information of the type associated with lexical formatives.

Presumably the additional information corresponds to the additional meanings of the ambiguity.

Finally, it is also assumed that the information accessed for each word is made available to a sentence integrator which determines sentence meaning by syntactic and semantic processing, or their equivalents (the process of evaluating the interaction of the appropriate deep structure for the string and the appropriate meaning for each of the lexical items).[4] The concerns of this paper center on how, when, and how much of, certain types of lexical information are made available to this sentence integration process.

General Statement of the Experimental Problem

Specifically, the focus of this paper is on the influence of particular lexical 'contexts' upon the interpretation and processing of other lexical items during sentence comprehension. The problems to be studied are those of 'when' a context is utilized during the time course of lexical comprehension (i.e., the locus of the effect) and just 'how' context has its effect. The 'when' question is one concerning the locus (in processing for comprehension) at which context influences the assignment of a meaning to a lexical item. The 'how' question is a matter of the psychological processes involved in such assignments; the mechanisms for the integration of information about two or more words during lexical processing. These two questions are interrelated--the former being a part of the latter. However, they can logically be separated, and will be studied and considered independently. The major concern here will be with the question of the locus of the context effect. The 'how' question will be elided for the moment, for several reasons. First, data on the 'locus' question can help to place some constraints on the plethora of possible models describing memory store and access. As the available data place few constraints on the possibilities for such models, delay seems warranted. Further, while several information processing models exist which incorporate mechanisms claiming to explain how context works (see, e.g., Norman, 1968; Morton, 1969), they are all extremely weak conceptually, with a distinct lack of explicitness and little or no supportive data. Thus, it seems best not to become involved in attempting to evaluate any one such model without more data to indicate what a good model might have to account for.

---

[4] Evidence exists (e.g., Foss, 1970) that, at least to some extent, semantic information is accessed and made available during the course of input of the sentence, and such processes do not wait until the end of the sentence. [However, see Hakes (1972), Cairns (1970) and Kimball (1974) for some thoughts as to what processes apparently do wait for completion of other processing.]

In considering the question of the locus of context effects, emphasis will be placed on the nature of prior context effects. It is possible to posit two general hypotheses concerning the effect of a prior lexical context upon a later-occurring lexical item. These will be called the Prior Effect Hypothesis and the Post Effect Hypothesis.[5] A Prior Effect would be a condition where the context predisposes a certain interpretation of a lexical item during language comprehension, decreasing the complexity of the comprehension task by limiting, narrowing, or biasing the processes which an interpretation is achieved, and doing so before the item is actually heard or accessed. A Prior Effect might occur, for example, if the processing mechanism were to place constraints on what pieces of information can be accessed for a particular item. As an illustration, were the word 'crane' to occur following a discussion of rare birds, one might expect the information associated with birds to have already activated the 'bird' interpretation of 'crane' such that upon hearing 'crane' access eventuates in the 'bird' interpretation without the 'heavy machinery' meaning ever being activated or considered.

---

[5] Foss and Jenkins (1973) have used the terminology of "Prior" and "Decision Point" in discussing these hypotheses. For reasons of conceptual clarity later in this paper, the latter term will not be used here.

A Post Effect of prior context would be one in which the context has some influence on comprehension of a lexical item only after that item has been fully processed (i.e., where access or retrieval of all semantic and other information has already occurred). This effect is similar to what one might expect when a pertinent context follows a lexical item; the later context aids in selecting some appropriate subset of the available information to retain and evaluate. Thus, the Post Effect Hypothesis suggests that even if the context is prior to the lexical item under consideration it can act on that item only after information about the item has been fully accessed. There currently exist some data which seem to support a Post Effect Hypothesis, data which will be reviewed shortly. However, the interesting question, and the focus of this research, is whether or not a Prior Effect does, in fact, ever occur. While the Prior Effect Hypothesis is the one which appears to agree most closely with one's intuitions about these matters, data supportive of this hypothesis are virtually nonexistent, and those data which are claimed to support the hypothesis are, in general, quite suspect. Antecedent to discussion of these data, however, it is necessary to consider the concept of ambiguity in language.

Rationale for the Study of Ambiguity

A large amount of research on comprehension has utilized and/or focused upon, ambiguity in language. Within a framework of transformational grammar, a sentence is considered to be ambiguous if the grammar relates its surface string to more than one semantic reading[6] and, in general, there are at least four types of ambiguity normally defined in linguistics: phonological/segmentation, lexical, surface structural (bracketing), and deep structural. As already indicated, the present paper is concerned with lexical ambiguity--the situation in which a word has more than one meaning.

The general interest in ambiguity among those who study language stems from many sources, but the basis for such interest in this paper is that lexical ambiguity provides a situation in which the information associated with the different readings of a word can be logically separated; this, in turn, provides a situation in which the effects of contexts upon information associated with the several readings of an ambiguity can be independently observed.

It might be felt that study of ambiguity is an investigation of only a special case of language processing--one not commonly used and therefore one which operates under other than 'normal' circumstances. However, the evidence suggests that ambiguity in language may quite possibly be the rule rather than the exception; rather than considering ambiguity as a special case in language processing, it might best be considered as the norm. To begin with, the reader may have noted that a great many lexical items can be, and are, ambiguous in the absence of any context. Further, Garrett (197) has pointed out that many everyday sentences are linguistically ambiguous but that other sentences sufficiently disambiguate them such that we are unaware of the ambiguity. Cairns (1970) has noted that, from a processing point of view, most sentences are ambiguous at some point prior to their completion. Finally, it seems that, given a sufficiently tortured context, a great many lexical items can take on a meaning other than the 'normal' one. [For example, Garrett (197) provides a situation where 'walking stick' is interpretable as an animate stick.] As interpretation of such a large amount of ambiguous input is a normal occurrence for the comprehension device, studies purposely focusing on the processing of certain ambiguities should directly reflect the nature of the general (normal) language processing mechanism.

---

[6] Traditionally in transformational grammar, ambiguity is regarded as a property of sentences (Katz & Postal, 1964). Thus, lexical ambiguities which have disambiguating contexts within the same sentence are linguistically unambiguous. This, unfortunately, does not solve a problem for the psycholinguist, but merely obscures it. Hakes (1971) has given a lucid account of why the sentence may not be a reasonable parameter in which the psychologist should view ambiguity and the processing of ambiguity.

There is some empirical evidence that not only are the meanings of an ambiguity separable but that they may also be stored separately in memory. It has been found that some ambiguous words (homographs) are judged to be words faster than are unambiguous words (Rubenstein, Garfield, & Millikan, 1970; Rubenstein, Lewis, & Rubenstein, 1971). This is not the case for all ambiguities, but only for those such as the word 'pipe'--called unsystematic ambiguities--where the two or more readings of the homograph are not conceptually related. On the other hand, ambiguities in which the readings are related in some fashion--called systematic ambiguities (e.g., plow)--were judged to be words no faster than unambiguous words. Rubenstein et al suggest these results indicate that unsystematically ambiguous words are represented more than once in the internal lexicon (presumably once for each different meaning). Thus, the reaction times are shorter to these words because, on the average, it is easier to find one of several possible readings for an ambiguous word than the single reading for an unambiguous word.[7]

Experimental Studies of Ambiguity

a) Effects of unbiased ambiguities on sentence processing

This interpretation of the storage characteristics of ambiguous words suggests an important question: How much information concerning an ambiguous word is actually available and used in normal sentence processing?[8] In the case of ambiguities, the question can be stated as whether information related to one or to several meanings is made available during processing (in unbiased contexts). Note that while listeners are rarely conscious of more than one meaning of an ambiguity in normal language processing, this may not be a true reflection of what occurs during processing; consciousness may well be quite removed from initial processing. (See Foss & Swinney, 1973, for further thoughts on this matter.)

---

[7] A recent study by Gough (personal communication) has shown the Rubenstein results to be difficult to replicate--although a trend in the appropriate direction was obtained.

[8] Foss and Jenkins (1973) have phrased this question in terms of how much information is transferred to a limited capacity working memory upon the occurrence of an ambiguous word.

From the one versus several readings question, the following hypotheses can be postulated: If the occurrence of an unsystematic lexical ambiguity causes several readings to be made available to the language processing device, this should result in greater immediate processing complexity (or processing load) than when unambiguous word occurs; however, if only one reading of the ambiguity is made available, then, all else being equal, processing complexity for ambiguous words should not differ from that for unambiguous words.

Some relevant studies of the effects of ambiguity upon processing complexity in unbiased sentences have been performed, and the evidence is that lexical ambiguities in sentences increase sentence processing difficulty. Foss (1970) presented lexically and structurally ambiguous sentences (and controls) to subjects who monitored for the initial phoneme of a word closely following the ambiguity (or the control).[9] In the basic phoneme monitoring task, monitoring reaction times were longer for the ambiguities than for their unambiguous controls. Further, the type of ambiguity had no effect on monitoring reaction times. This was taken as evidence that more than one reading for each of the ambiguities was accessed, causing a subsequent increase in processing difficulty.

---

[9] In a phoneme monitoring task, the subject is instructed to comprehend a sentence, and at the same time, listen for a particular (priorly specified) phoneme beginning some word in the sentence. RT's to respond to the phoneme are recorded. The premise governing use of this task is that the central processing mechanism used to deal with comprehending sentential material overlaps that used for the phoneme search, and, at points where sentential processing is more difficult, phoneme monitoring latency should increase. This premise has been justified in many experimental situations to date. (See, e.g., Foss & Cairns, 1970; Foss, 1970; Foss & Lynch, 1967).

It seems that several hypotheses can be offer   to account for the increased processing difficulty encountered in the presence of lexical ambiguities. The effect could be due to the time necessary to 'look up' o access several separately stored readings; or, it could be due to the absolute amount of information (perhaps all stored in the 'same' place) that must be accessed (or, alternatively, transferred to a working memory); or, it might be due to the effects of a decision process which _chooses_ the most appropriate reading to be made available to the sentence integration device. These hypotheses will be considered in greater detail below.

There are other data consistent with the Foss finding. Garrett (1970) found that subjective location of a 'click' (which was actually placed on an ambiguous word) was displaced further when the sentence was disambiguated only _after_ occurrence of the ambiguity, than when disambiguation occurred in a clause prior to the ambiguity. In accord with the general interpretation given 'click' studies, this suggests that unbiased ambiguities are areas of increased processing complexity. In all, then, lexical ambiguities appear to increase processing complexity during sentence comprehension.

There is also evidence (Foss, Bever, & Silver, 1968; Cairns, 1970) which suggests that the increased processing complexity effect for lexical ambiguities is reasonably short lived. The relevant data are from studies which utilized both lexically and structurally ambiguous sentences. Subjects were presented with a sentence, followed by a picture (or, in the Cairns study, by another sentence). Subjects were then required to decide whether the picture corresponded to the meaning of the sentence. The 'related' pictures were sketches of activities outlined in the sentence; one set of pictures depicted the more likely (expected) interpretation of the sentence, and another set depicted the less likely, but acceptable, (unexpected) interpretation. Likelihood of interpretation was determined beforehand. The noncorresponding (incorrect) pictures differed from the sentence with respect to the subject, object, or verb. For structural ambiguities there was no significant difference in verification time between the expected and unexpected pictures (relative to their controls). However, it was found that verification of unexpected pictures for lexical ambiguities took significantly longer than for expected pictures (again relative to their controls). This has been interpreted as indicating that decisions concerning lexical ambiguities are made before the picture has been interpreted (causing necessary reprocessing in the case where the unexpected picture occurs), but that both structures are carried along and a decision not made between them before the picture has been interpreted (sometime after the end of the sentence) for structural ambiguities. (Or, at least, a decision may be made in favor of one, but the other structure is still available.) Even more to the point of the duration of the lexical ambiguity effect is a study by Cairns (in preparation) in which she used a phoneme monitor task where the monitor target was placed either immediately after an unbiased lexical ambiguity or two words after the ambiguity. While she obtained the Foss (1970) effect (increased processing in presence of the ambiguity) with the 'immediate' placement of the monitor target, the condition where the target was two words removed from the ambiguity produced no ambiguity effect. Thus, it appears likely that the question of the appropriate reading for a lexical ambiguity (presumably the source of the increased processing effect) is resolved fairly quickly.

Experimental Studies of Ambiguity

b) Effects of post-disambiguating context upon processing of ambiguities in sentences

Data from experiments dealing with ambiguities and post-disambiguating contexts also provide some evidence of increased processing complexity in the presence of lexical ambiguities, and suggest that multiple readings of the ambiguity appear to be available for a short time. There are several pieces of evidence for this. Caplan (1971) ran an experiment in which subjects listened to sentences and watched for a word which was visually presented on a screen in front of them. Subjects had the task of pressing a button to indicate if the word they saw had occurred in the sentence, or not. (This is called a probe-latency experiment.) Subjects heard sentences which contained an ambiguity (or an unambiguous control word) followed later in the sentence by a disambiguating context. The probe word was either the ambiguity or the control word, and it was presented on the screen either before the post-disambiguating context (but after the ambiguity) or after the disambiguating context. It was found that when the probe occurred prior to the disambiguating context, reaction times to the probe were faster for the ambiguous than for the control conditions. There was no significant ambiguity-control difference when the probe followed the disambiguating context. Assuming that multiple representation facilitated recognition of a probe, this data supports the interpretation that while one reading had been chosen following the post-context, both readings were available prior to that time.

Lackner and Garrett (1972) had subjects perform a paraphrase task in a dichotic listening situation while attending to only one ear (which contained an ambiguous sentence). A disambiguating sentential context was simultaneously presented in the other ear, but at 5dB below the level of input to the attended ear. In this experiment, it was found that for 14 of the 45 sentences, the disambiguating 'context' in the unattended ear occurred after the ambiguity was presented in the attended ear. Of these 14, 13 of the paraphrase interpretations were made in the direction of the reading favored by the biasing context. This effect was significant. (It should be noted that subjects could not report what the context word was; apparently it was masked from conscious interpretation.) Lackner and Garrett conclude that this supports a theory that multiple readings for a lexical ambiguity are computed upon its occurrence. (Otherwise, they argue, the unattended post-context should not have had any effect--i.e., it should have been treated as unrelated.) In a similar study, MacKay (1973) presented unbiased lexically ambiguous sentences to one ear (to be attended by subjects) and a disambiguating word, simultaneously, to the other (unattended ear). Subjects were required to shadow the attended sentence and, at the end, indicate which interpretation of the ambiguity they heard. MacKay found a significant shift toward the context even though subjects were not aware of the nature of the unattended words. This, again, suggests that both readings of an ambiguity are available, at least momentarily.

While these data fit with a general hypothesis that all readings for a lexical ambiguity are generated and held for a decision, Hakes (1971) points out that there are limits to the length of time such a decision can be withheld in hopes of disambiguation.[10] The anecdotal evidence from garden path sentences strongly supports this notion. For example, listeners generally report that they begin to interpret #1 as "The old man searched for his eyeglasses," and then later reinterpret it when later information (the last clause) is heard.

> #1. The old man searched diligently around the dimly lit room for his glasses and finally found them on the table near the window, filled with sherry.

Thus, while post-disambiguation can aid the comprehension device in choosing one of the several readings of an ambiguity as relevant, a decision is made reasonably quickly without its use when it does not occur in close enough proximity to the ambiguity. Again, recall that the evidence from the Cairns (in preparation) study suggests that lexical ambiguities may be resolved by the time the next two words following it are processed.

Studies of Sentential Processing of Ambiguities Involving Prior Context

The evidence from studies using prior-disambiguating contexts is more difficult to evaluate than that from post-disambiguation experiments, and the few studies relevant to the nature of prior disambiguation effects have not proved even partially conclusive. Probably the clearest experimental evidence in support of a Prior Effect Hypothesis is found in a series of experiments by Meyer and his associates (Meyer, 1971; Meyer & Schvaneveldt, 1971; Meyer, Schvaneveldt, & Ruddy, 1972). It is worth noting that Meyer's experimental task was not one which required normal sentence comprehension. The basic task is one in which subjects are required to decide if strings of letters are English words, or not. It was found that when two strings are presented simultaneously positive decisions for associated words (Doctor-Nurse) are facilitated relative to decisions for unassociated words (Butter-Nurse). Further, this facilitation occurred even when the associated words were separated by the processing of an unrelated word. Meyer accounted for these effects in terms of a 'spreading activation' model of semantic memory access. Such an account holds that words (or their equivalent) are stored in such a way that items of similar semantic content are (metaphorically) near each other. 'Activation' is seen as spreading to 'nearby' words from the one already activated. Thus, a Prior Effect Hypothesis is at least partially supported by the Meyer data, in that activation of one word apparently raises the level of activation of another (regardless of whether that other word is ever called for) making potential access to the later word somehow easier.

---

[10] Also see text, earlier, for comments on the Foss, Bever, & Silver (1968), and Cairns (1970) results which related to the question of how long both readings of a lexical ambiguity are available during sentence processing.

Other experiments studying prior disambiguation have used sentence comprehension paradigms. Garrett (1970) performed an experiment in which subjects were to specify the location of a click (which was placed on a lexical ambiguity in the sentence) while they listened to sentences. The hypothesis behind this task is that click location accuracy is a function of how much work the processing device is having to perform; the more complex the processing, the larger the expected error in click location. Garrett's findings (as mentioned earlier) were that predisambiguation conditions had more accurate click placement than did post-disambiguation conditions. This would tend to support a Prior Effect Hypothesis (as there was apparently reduced processing complexity for the priorly disambiguated ambiguity). There is some question, however, as to whether this can be considered to be a prior effect of context. In this experiment, subjects indicated the click location (by marking a written transcript) only after the entire sentence was heard. There is little guarantee that the placement of a line on a written transcript by a subject, in the place where he thought he heard the 'click', will hear any strong relation to the on-line processing difficulty he experienced while attempting to process it. Thus, one must be extremely careful and skeptical in drawing conclusions from such data; there are far too many variables at work (i.e., memory) to be confident that the effect reflects a 'prior' reduction in the momentary processing load occasioned by the presence of an ambiguity.

Conrad (1974) used a prior-context design to investigate lexical 'activation' in a sentential framework. Auditory presentation of a single sentence to subjects was immediately followed by the visual presentation of a word printed in colored ink. Reaction times to name the color of the ink the word was printed in were recorded. Subjects then had to recall the sentence (aloud). The color naming task (Stroop test) was assumed to be a test of interference: the higher the current level of 'activation' of the semantic meaning of the visually presented colored word, the more it should interfere with color naming. In Conrad's experiment, the visually presented word was either: 1) an ambiguous word which had occurred in the sentence just presented (e.g., fork), 2) a 'category name' (e.g., utensil) related to the reading of the ambiguous word toward which it had been biased by a prior context (appropriate), or 3) a category name (e.g., path) toward which the ambiguous word had not been biased (inappropriate). Conrad ran several experiments, but the relevant results involve sentences where, in the experimental condition, the last word in the sentence was an ambiguous word. Each ambiguous word was used in 12 different sentences. In half of these sentences, the ambiguous word was disambiguated by prior context (three sentences in the direction of each of the two readings). The other six sentences were left ambiguous. The results were that the sentences containing ambiguities had longer reaction times than appropriate control sentences, and there was no significant difference between reaction times to the three different word targets in the experimental sentences which had been disambiguated. Conrad took these findings to indicate that the ambiguous word, when heard in the sentences, activated both readings related to it, regardless of whether it had previously been disambiguated or not (i.e., interference was equal for both appropriate and inappropriate category names, thus the assumption is that both were activated). Note that a Prior Effect model would have been supported only if just the appropriate category name words had resulted in more interference, and the inappropriate category name words had not. It is very important to note that in this experiment only eight different ambiguous words were used and that all subjects heard each ambiguity

12 times (i.e., in 12 different sentences). It should not have taken the subjects long to notice the presence of these ambiguities, and to question themselves concerning each of them--which, in Conrad's terms, would have 'activated' both related readings. Thus, her results are far from conclusive.

There are data that relate somewhat more directly to the on-line processing of ambiguities in predisambiguated contexts. These are from a series of phoneme monitoring experiments performed by Foss and his associates. Foss and Jenkins (1973) performed an experiment in which one-half of the sentences contained an ambiguous word and the other half contained an unambiguous word that was synonymous with one of the readings of the ambiguity (e.g., stock-cattle). Orthogonally, one-half of the sentences had a disambiguating context word prior to the ambiguity that was biased (as judged by the experimenters) towards one reading of the ambiguity. (Foss and Jenkins note that the neutral context words were equally related to both readings of the ambiguities, and that the biased contexts did not totally rule out the 'other' readings of the ambiguities, but, rather, made them less appropriate). Both readings of each ambiguity were tested in this fashion. The monitor target was never more than two words after the ambiguity (or its control). In this design, a Prior Effect Hypothesis predicts an interaction between the ambiguity and context variables. (That is, reaction times following ambiguous words should be longer than those following unambiguous words in the neutral context conditions, but this difference should disappear with the biased context if it acts in a truly 'prior' fashion.) The results were that reaction times to ambiguous sentences were significantly longer than those to unambiguous sentences, but the predicted interaction, while in the predicted direction, was not significant. Thus, the Prior Effect Hypothesis did not find support here, although the Post Effect Hypothesis did.

Foss and Hassler (personal communication) performed a monitoring experiment where an ambiguous noun (or a control) was immediately preceded by a disambiguating adjective. (The adjective was omitted in the control condition.) The target word closely followed the noun. The appropriate interaction was, again, not significant, but was found to be in the direction predicted by a Prior Effect model.

Subsequently, Cutler and Foss (1974) performed a phoneme monitoring experiment that expanded upon Foss and Hassler's in that either none, one, or two adjectives preceded the ambiguous (or control) noun in the sentence. The design also included a variable testing the effect of the position of the ambiguity in the sentence; ambiguities occurred, factorially with the other variables, in both subject and object positions. The results of the experiment showed significant ambiguity and context main effects, as well as a significant context by position interaction. However, the ambiguity by context interaction was not significant and thus, again, Cutler and Foss concluded that a Prior Effect Hypothesis was not supported. (There was, however, a trend in the appropriate direction for the object position data only.)

The sentence-processing experiments to date, then, have failed with amazing persistence to find a Prior Effect of context. It might seem appropriate at this point to conclude that there is no such effect, but for a couple of very important points: For one, it appears to this writer that many of the crucial tests involving context and sentence processing have not been performed. Further, many of the experiments with negative results currently in the literature have included major flaws in materials or methodology. The importance of having a definitive answer to the question concerning existence of a Prior Effect necessitates more careful study of the problem. The experiments proposed here, then, are an attempt at a more careful test of whether, in fact, a Prior Effect may occur. To elucidate the approach, it is necessary to spell out a few of the criticisms of previous research in more depth.

Discussion and Evaluation of Methodology and Materials Used in Studies of Prior Context and Ambiguities

To begin with, a question can be raised as to what logically constitutes a disambiguating context. That is, what constitutes the possible set of disambiguating context materials for a hypothesized Prior Effect model?

There appear to be two general (potential) aspects of disambiguating relationships which hold between a prior context and an ambiguity. One aspect is the a posteriori ability of the context to 'choose' a meaning for the ambiguity. This is the ability of a context to effectively distinguish between two or more readings of an ambiguity (after all the readings are known) by being more related to one and, more importantly, by being able to eliminate the 'other' reading(s) on some basis (such as conflicting semantic relationships, etc.). The second aspect of a disambiguating relationship is the a priori strength of the prior context-ambiguity relationship; the semantic or associative strength of the prior context with one reading of the ambiguity. For example, this might be thought of as the likelihood that context word X will have one reading of the ambiguity Y high on its own list of "closely related words and concepts."

The materials used in previous experiments include samples of both weak (or nonexistent) and strong members of the a priori and a posteriori disambiguation categories. For example, Cutler and Foss (1974) used the context "black" to disambiguate the ambiguity "glasses." If one considers the set of words or concepts highly associated with the word "black," it does not appear likely that "glasses" will enjoy a prominent place in that set. Thus, this constitutes a weak a priori relationship. Similarly, "black" does not appear to strongly disambiguate "glasses" even in the a posteriori sense; that is, both readings of "glasses" are acceptable in the presence of this context. In all, perusal of the materials in previous 'prior disambiguation' experiments reveals relationships which fall into all four cells of a weak-strong X a posteriori-a priori analysis. For example, "black-glasses" constitutes both a weak a priori and a weak a posteriori relationship; "wire-rimmed glasses" appears to have a weak a posteriori, but a strong a priori relationship; "gas-pipe" has a strong a posteriori, but a weak a priori relationship;

and finally, the example "baseball-bat" has both strong a priori and a posteriori relationships.

It seems reasonable to conclude that, while either strong a posteriori or strong a priori context-ambiguity relationships could produce results supportive of a Post Effect Hypothesis, only strong a priori relationships can be expected to produce a true Prior Effect of context. As a great many of the materials used in previously mentioned experiments have not provided strong a priori relationships, it seems that these previous results must not be considered definitive and that the context-ambiguity relationship must be more carefully considered and controlled in studies of prior context effects.

The position of the context relative to the ambiguous word during sentence comprehension is also of importance in such investigations, and it is a variable that has not been considered in previous work. If, for example, the context is placed too closely in front of the ambiguity, there may not have been time for it to be fully processed by the time the ambiguity is beginning to be processed; thus it would be incapable of producing a Prior Effect. This is a simple point but many of the materials in the earlier experiments (see, in particular, Cutler & Foss, 1974) which rejected the Prior Effect Hypothesis are of this design. At the other end of the continuum, while it might beem likely that there are limits to the time or distance over which a prior context can be effective, there is little empirical evidence upon which to define such limits. One piece of relevant data is found in the work of Meyer et al

(1972). Meyer et al found relative facilitation for making a word/nonword judgment in the presence of an associated word, compared to an unassociated context control. It was found, further, that this facilitation tended to decay as a function of the delay between presentation of the associated words and that this function asymptoted at a maximum of about four seconds (although the facilitation did not disappear completely).* There was no task required of the subjects in the intervening four seconds; thus it is possible that some form of rehearsal occurred, although Meyer et al did not believe this to be the case. Regardless of the validity of Meyer's four second value when applied to sentential processing, the general hypothesis of 'decay of activation' appears to be empirically testable, and, in fact, study of context-ambiguity distance appears required before rejection of the Prior Effect Hypothesis can be considered conclusive.

A related problem to consider is that of whether(and how)the structure of the sentences (involving such variables as clause structure and clause boundaries) may affect the context-ambiguity interaction. Evidence has been accumulating for some time concerning the potential importance of clauses in language processing. Bever, Lackner, and Kirk (1969) and others (see Garrett, 1970) have investigated errors in locating clicks in sentences. The general results of these experiments are that subjects tend to misplace clicks toward the major sentential (often embedded clause) boundaries. This has been interpreted as indicating that clauses are cohesive units of processing in language and that clicks are displaced perceptually out of these areas into places of lesser processing difficulty--

the clause boundaries.[12]

Further, Caplan (1971), using the probe-latency technique, found that reaction times to a probe occurring after the end of a sentence were faster when the probed-for word had occurred in the sentence's last clause than when it had occurred in an earlier clause (distance was constant in this design). This suggests that information in the previous clause had been interpreted and removed from immediate processing (working memory) whereas the last clause was still available in the working memory. Jarvella and Herman (1972) in a memory recall task, also found evidence that clauses (and not sentences) are interpreted as 'whole' units in psychological processing. Bever, Garrett, and Hurtig (1973) in a sentence completion experiment found that reaction times to structural ambiguities in incomplete sentences, but completed clauses, were facilitated (compared to nonambiguous controls) relative to similar materials ending in incompleted clauses. Again, this is evidence that the clause plays a major role in sentence processing. The exact nature of this role is not at all determined at this time. Similarly, the nature of the processes which occur during the comprehension of clauses is not known; the cohesive properties of the clause could be due to the active process that was earlier called 'sentence integration' (perhaps triggered when the processor decides the end of a clause has been reached) or they could be a function of a completed clause being stored as a unit after being removed from working memory. Regardless of the actual nature of the processes involved, the clause appears to be a major unit of processing during sentence comprehension. As such, it may be an important variable to consider in studying prior context effects upon ambiguities.

---

[12] Note that an alternate hypothesis (see Hakes, 1971) is that final integration of clausal information is performed at a clause boundary, and, hence, perception of the 'click' takes place where such decisions are made.

It is possible, for example, that both context and ambiguity must be in the same clause in order for one to have an effect upon the other during active processing. If, for example, the results from the processing of a completed clause are shifted from immediate working memory, a context in a clause preceding the ambiguity might never be brought to bear on the ambiguity. Alternatively, one of several models of lexical access and clause processing suggested by Hakes (1972) holds that lexical decisions may not be made in their entirety until the end of the clause. In this model, a context within the same clause as the ambiguity would not yet have been fully processed, and thus might not be expected to have an effect upon the ambiguity. In fact, under this model, a lexical context would have to be in a clause prior to the ambiguity to be fully available to affect processing of the ambiguity.

In order to investigate these possibilities a re-analysis of the Foss and Jenkins (1973) data was performed. Their materials were re-organized on the basis of whether the prior, disambiguiating context and ambiguity were in different clauses or were in the same clause. It was found that just in those cases where a clause boundary intervened between context and ambiguity, prior disambiguation tended to reduce differences between ambiguity and control conditions—a finding supportive of a Prior Effect Hypothesis. This trend was not present when the context and ambiguity were in the same clause. The cell entries for this re-analysis could not be well balanced, or statistically tested in any meaningful way, and thus such ad hoc re-analysis can only be considered suggestive, at best.

However, it appears important to consider the clause boundary variable in future tests of the Prior Effect Hypothesis.

A pilot study was run in an attempt to investigate the clause boundary variable. This experiment, utilizing the phoneme monitor task, also attempted to study the variable of context-ambiguity distance. The results of an analysis of variance on data from this experiment yielded no significant effects, primarily due to variance attributable to materials. The study was not a total loss, however, for post hoc examination of the materials revealed a fact which had been missed earlier. When there was a clause boundary (actually, a sentential boundary) separating the context and the ambiguity, the ambiguity always occurred in the 'subject' position in its sentence. When no clause boundary separated the context and ambiguity, the ambiguity was in the 'object' position in the sentence. Further, just in the situation where the ambiguity was in the object position, a 'distant' context (14-25 syllables prior to the ambiguity) produced a relatively greater facilitation in processing the ambiguity than did either an immediately prior context, or no context at all. Given this trend and the related subject-object trend noted in Cutler and Foss (1974), it appears that the relative position of the test item ('subject' or 'object') might be an important variable to consider in the study of context-ambiguity interaction.

Finally, it should be noted that in phoneme monitoring experiments the placement of the monitor target relative to the item to be studied (in this case, the ambiguity) is very important to consider. As cited above, Cairns (in preparation) reports an ambiguity study using the phoneme monitor task where she placed the monitor target either immediately after an unbiased lexical ambiguity, or two words after the ambiguity. While she obtained the Foss (1970) effect (increased processing in presence of the ambiguity) with the 'immediate' placement of the monitor target, the condition where the target was two words removed from the ambiguity produced no ambiguity effect. While an effort is normally made to place the monitor target as close to the area of interest as possible, the Cairns study further points up the necessity for care in this regard. Related to this, the pilot study reported above also found that the ambiguity effect is not clearly obtained when a clause boundary intervenes between the ambiguity and the monitor target, even if the target is the word immediately following the ambiguity.

In summary, there are a number of materials variables and theoretical points which have not been adequately controlled or examined in previous attempts to study the effects of prior context upon the processing of lexical ambiguities. The present experiment was specifically designed to study the effect of prior context upon the processing of lexical ambiguity in a sentence framework, with the particular goal of maximizing the possibility for achieving a Prior Effect of context. The experiment utilizes the phoneme monitor task and the conceptual design is basically the same as that for the Foss experiments; the differences between this and previous experiments can largely be described in terms of the use of the strongest possible context-ambiguity relationships, consideration of the clause boundary and distance variables mentioned above, and careful control of possible disruptive or confounding conditions.

CHAPTER II

METHOD

Design and materials

The materials for this experiment consisted of thirty sets of basic sentence-pairs (two sentences, one immediately following the other). Each basic sentence-pair had six variations. These six variations constituted a set. The two sentences of any sentence-pair did not constitute an anomalous utterance when heard together. Differences among the six basic sentence variations were characterized by two factorially combined variables: Test Item (TI) and Prior Context (C).

The TI variable contained two conditions: a) presence of an ambiguous word, or b) presence of a control word which was roughly synonymous with one reading of the ambiguous word with which it was matched. The TI words were all nouns and appeared in the direct object (i.e., in the predicate phrase) of the second sentence of the sentence-pairs. The words which composed the ambiguous and control categories of the TI conditions were equated for frequency and length in syllables (see Appendix A for these values).

The ambiguous words were judged to be 'unsystematically ambiguous' (using the Rubenstein definition, as given above) on the basis of the experimenter's intuition. These decisions were confirmed by the intuitions of two other psycholinguists. Ideally, control words in this study would be completely unambiguous. Such an ideal is, however, difficult, if not impossible, to attain in practice. There are relatively few words in the English language which, taken completely out of context, are completely unambiguous. Consequently, the control words selected were often themselves ambiguous. They differed from the ambiguous words in that the source of ambiguity for which the latter were chosen was not present for the former. In this sense, the control words were less ambiguous than the ambiguous words.

It is, of course, difficult to determine a priori whether this difference in amount of ambiguity actually obtained. The test of whether it did is whether, in the absence of a prior disambiguating context, phoneme monitoring reaction time to the ambiguous words is longer than to the control words. If this result obtains, we may take it as prima facie evidence that the control words are at least less ambiguous than their ambiguous counterparts.

The Prior Context (C) variable contained three conditions: a) presence of a relevant context word (or words) in the second member of a sentence-pair, 1-3 syllables prior to the TI (the Immediate Context condition; b) presence of this same relevant context in the first member of the sentence-pair, 10-40 syllables before the TI (the Distant Context condition); c) absence of a context that related to the TI in the manner of the Immediate or Distant contexts. The context was the same, or nearly so, for the Immediate and Distant conditions.[1] Only a single context

---

[1] The Distant Context was always in the first sentence of the pair and, hence, across a sentential boundary from the TI. The Immediate Context was in the second sentence and in the same clause as the TI. Thus distance and clause boundary effects were confounded in this arrangement.

(or no context at all) appeared in any one sentence variable. All contexts words were judged by the experimenter and by one other psycholinguist to fall into the strong a priori and a posteriori classifications, whenever possible, in terms of their relationship to one reading of the ambiguity.

The word which contained the phoneme monitor target was always in the same clause (but not necessarily the same phrase) as the TI and was never more than two words after the TI. The monitor target was the same for all variations of a basic sentence-pair. The target words did, however, vary across the 30 different sets of basic sentence pairs (see Appendix B for a list of the experimental sentences in schematized form; the monitor targets are indicated there by the symbol "Δ"). The following example illustrates one set of basic sentence-pairs, where N = No context, I = Immediate Context, D = Distant context, A = Ambiguity, U = Unambiguous control. Thus, for example, NA stands for the variation with no context and an ambiguity.

1. (NA) Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several bugs in the corner of his room.
Δ

2. (NU) Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several insects in the corner of his room.
Δ

3. (IA) Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches and other bugs in the corner of his room.
Δ

4. (IU) Rumor had it that, for years, the government
building had been plagued with problems. The man
was not surprised when he found several spiders,
roaches and other insects in the corner of his room.
Δ

5. (DA) Rumor had it that for years the government
building had been plagued with spiders, roaches
and other problems. The man was not surprised
when he found several bugs in the corner of
his room. Δ

6. (DU) Rumor had it that for years the government
building had been plagued with spider, roaches
and other problems. The man was not surprised
when he found several insects in the corner of
his room. Δ

Six tapes were made from these materials. Each tape contained only one variation of each basic sentence-pair, and the six types of variations were equally represented on each tape. Assignment of a particular sentence-pair variation to a tape was accomplished by an ordered roll-over technique. Each subject heard only one tape. All tapes contained identical filler sentence-pairs of two types, randomly interspersed among the test sentence-pairs. One type consisted of 20 sentence-pairs in which the monitor target occurred earlier in the second sentence of the sentence-pair[2] than it did for the experimental sentence-pairs. The purpose of this type of filler was to dissuade subjects from believing that the target would always occur late in the second sentence. (Appendix C lists these filler sentence-pairs). The second type of filler consisted of 20 sentence-pairs which contained no monitor target. These were included to avoid an increase in expectancy for occurrence of a monitor target near the end of the sentence-pair in those cases where it did not occur earlier (Appendix D lists these filler sentence-pairs). There were no systematically varied ambiguities in any of the filler sentence-pairs.

---

[2] Generally, the monitor occurred in the subject or verb position in these filler sentences.

A signal was placed coincident with the onset of each phoneme monitor target on a separate track of the tape. It was used to activate a timing mechanism. Exact placement of the signal was determined by moving the tape back and forth manually across the tape heard until it was judged that the onset of the target sound had been reached. This signal could not be heard by subjects listening to the sentences.

In sum, each tape consisted of 40 filler sentence-pairs and 30 experimental sentence-pairs (which could be classified into six conditions). These were arranged in a mixed between-within design, where there were two within-subjects variables (Test Item and Context) and one between-subjects variable (Groups). The between-subjects variable was the result of counter-balancing for materials and presentation across subjects (i.e., assignment of sentence-pairs to tape conditions).

Subjects

The subjects were 90 undergraduates from The University of Texas at Austin. They participated for 45 minutes in partial fulfillment of a psychology course requirement. Fifteen subjects were randomly assigned to each tape condition. As there was a potential for five scores per condition for each subject, an a priori decision was made to eliminate any subject from the analysis who failed to yield at least two scores per condition for any of the six conditions. This was felt to be justified because all analyses were to be based on means for each condition. On this basis, data from eight subjects were eliminated; eight other subjects participated in their places.

Procedure

Up to four subjects were tested at a time. Each subject sat in an isolation booth and could see neither the experimenter not the other subjects. Subjects listened to the 70 binaurally presented sentences over headphones. The sequence of events on each trial was as follows: Each trial began with the number of that trial, followed by specification of a phoneme monitor target for which subjects were to listen, followed by the word "ready," followed by a sentence-pair. Subjects had two tasks. One was to listen for the target phoneme and to press a button (on which they rested their preferred hand at all times) as soon as they heard the specified target sound in each sentence. Subjects were instructed that the target would occur only in the second sentence, if at all, and that it would only be at the beginning of a word.

The subjects were also told to listen carefully to both sentences of each sentence-pair, and, after each trial, to indicate on a five-point scale how closely related they felt the two sentences were. Subjects were provided with a response sheet on which they could place a slash in the appropriate slot (see Appendix E for a sample response sheet). This task was included to help insure that subjects would attend carefully to the sentences. (See Appendix F for a copy of the instructions which were read to the subjects.)

At the end of the session, subjects were questioned concerning whether or not they were aware that some of the sentences had contained ambiguities. This task was included so that performance of subjects who were aware of the presence of ambiguities could be studied separately from that of ones who were not.

Timing and storage of all button-press reaction times was controlled by a PDP 8/I computer. Timing was initiated by the unheard signal which had been placed coincident with the onset of the phoneme target; it was terminated by the subject's button push. Reaction times longer than 2200 milliseconds were excluded. In previous experimentation using the monitor task, it has appeared as though reaction times longer than this often reflect a reprocessing of the input or a response to an incorrect target.[3] In addition, presentation of the sentence was quite often completed in less than two seconds following occurrence of the target phoneme. Processing which occurs after the sentence is completed may well be of a different nature than that occurring during ongoing comprehension.[4] For those reasons, the computer was set to allow only 2200 milliseconds for a response before presenting the subsequent trial.

---

[3] See Foss and Jenkins (1973) for further discussion along these lines.

[4] In order to discover, post hoc, if imposing a limit (such as the 2200 msec ceiling, here) would seriously alter the results, several tests were made. An analysis identical to those which will be discussed below was performed on data using only scores no longer than 1500 msec. (This required elimination of only 24 scores out of a possible 2700—less than 1%—which gives some idea of how compact the distribution was). The statistical results were identical to those which will be reported below for scores up to a ceiling of 2200 msec.

It should also be noted that the total number of omitted responses, whether due to the ceiling cut-off or merely to a subject not hearing a target, was nearly 10% of the total responses. Of the scores available for analysis, the mean reaction time was 525 milliseconds and the range was 85-2010 milliseconds.

CHAPTER III

RESULTS

The results which are of major interest in this study involve comparisons of the relative effects of processing ambiguities in the condition containing no context with those same effects for the two conditions containing biased contexts. Recall that the Prior Effect Hypothesis predicts that any increase in processing complexity occasioned by the presence of an ambiguity (compared to an unambiguous control) in the No context condition should disappear in the presence of an appropriately biased context. In the present experiment there are _two_ context conditions. While it may be that both context conditions will produce results which support the Prior Effect Hypothesis, it may also be that only _one_ of the context conditions will provide data which support this effect. (Refer to the Introduction of this paper for justification for this statement.) Thus, while a significant Ambiguity by Context interaction resulting from statistical analysis involving all 3 context conditions is reasonable evidence in support of Prior Effect Hypothesis, lack of this interaction does not disprove it; individual examination of each of the context conditions is necessary to determine whether or not a Prior Effect Hypothesis is to be supported under any condition. On the other hand, the Post Effect Hypothesis predicts no difference in the ambiguity processing load effect between the No context condition and either of the biased context conditions.

Six mean reaction times were computed for each subject, one for each of the six conditions of the Context by Ambiguity interaction. Table 1 displays total mean reaction times for these conditions, collapsed

CONTEXT

	None	Immediate	Distant
Ambiguous words	556	527	517
Nonambiguous words	514	540	495

Table 1. Mean reaction times, in milliseconds, to the six conditions of the ambiguity by context interaction.

across Groups. It can be seen that, in the condition with no biasing context, mean reaction times for sentences containing ambiguities are longer than those for unambiguous controls. Further, this difference is far larger than the ambiguity-control difference in either of the two context conditions. Informally, then, the data look as though they may conform to the predictions of the Prior Effect Hypothesis.

In order to test the hypotheses, the mean reaction times were submitted to an analysis of variance with Ambiguity, Context, and Group conditions as major variables. (Subjects were nested in the Groups condition.) See Appendix G for the source table for this analysis.

There was a significant main effect for the Context variable, $F(2,168) = 5.424$, $p < .01$, indicating that the reaction times differed significantly across the three context conditions, a finding that might well be expected given the basic hypotheses of this experiment. The main effect for Ambiguity was not significant, $F(1,84) = 3.478$, N.S. Most importantly, the Ambiguity by Context interaction was significant, $F(2,168) = 4.083$, $p < .025$. In order to examine this effect more fully, planned t-tests were run on the relevant comparisons. A t-test comparing data from sentences containing ambiguities with data from their unambiguous controls in the No context condition revealed the observed differences to be significant, $t(89) = 2.30$, $p < .025$, indicating that the ambiguity effect obtained in earlier experiments was also obtained here. Among other things, this difference provides evidence that the ambiguous and control words used in this experiment differed in ambiguity as intended. Another planned t-test indicated that, in the condition containing a context "Immediately" prior to the Test Item, reaction times for ambiguous and unambiguous-control conditions did not differ significantly, $t(89) = -0.203$, N.S. Similarly, the reaction times for ambiguous and control conditions did not differ significantly in the presence of a Distant context, $\underline{t}(89) = 1.57$, N.S. In sum, the presence of an ambiguity in the condition containing no relevant context led to longer reaction times (evidence of increased processing complexity) than an unambiguous word in the same condition. However, there was no difference in reaction time between ambiguous and unambiguous control conditions in the presence of a relevant context for either the Immediate or the Distant context conditions.[1]

Returning to consideration of the results of the analysis of variance, the main effect for Groups was not significant, $\underline{F}(5,84) = 1.852$, N.S. However, the Groups variable was involved in two significant interactions: Groups by Ambiguity, $\underline{F}(5,84) = 9.088$, $\underline{p} < .01$; and Groups by Context by Ambiguity, $\underline{F}(19,168) = 5.036$, $\underline{p} < .01$. As the above-mentioned Context by Ambiguity interaction is an important factor in this experiment, it seems necessary to comment on the significant Groups by Context by Ambiguity interaction. It is difficult to determine the logical significance of this effect because, while the Context by Ambiguity interaction does differ in magnitude and, to some extent, in direction across the different Groups conditions, comparison of these different Context by Ambiguity interactions across groups compares reaction times obtained from different sets of sentences. It is certainly a reasonable expectation that there will be variance in response latencies across sentences, or small subgroups of sentences. Furthermore, comparisons of Context by Ambiguity interactions across Groups involve comparing reaction times to different phoneme monitor targets. Reaction times to respond to different phoneme targets have been shown to differ drastically (see Foss & Swinney, 1973; Savin & Bever, 1970); thus, on the bases of these sources of variance it seems likely that relationships within the six Context by Ambiguity conditions might well be expected to differ across groups.

---

[1] These same comparisons were also made with a more conservative test, for, while the planned orthogonal t-tests are justified, and while they do provide a strong test of whether the ambiguity effect disappears in the context conditions, it might be felt that the significant effects found in the No context condition are suspect. Thus, a Neuman-Keuls test was run on the Ambiguity by Context interaction. This test revealed a significant ambiguity-control difference for the No context condition ($\underline{p}<.01$) but did not find any other (meaningful) effects to be significant. Thus, the effects reported above for the Context by Ambiguity interaction appear to be robust. See Appendix H for a summary of these findings.

A strong argument could be made that these potential sources of variance are the main sources of variance in the triple interaction, if the overall magnitude and direction of the Context by Ambiguity interaction is shown to hold up for sentences in general. In order to determine if this was the case, means were calculated for the six conditions of the Context by Ambiguity interaction for each sentence. An ad hoc Wilcoxen Matched-Pairs, Signed-Ranks Test, which compares both magnitude and direction of differences, was performed on these means taken across all sentences.[2] Means for Ambiguous and Nonambiguous control conditions were compared for each of the three context conditions. For both the Immediate and Distant context conditions there was no significant difference between ambiguous and nonambiguous control conditions, $z = 0.415$, 1.541, respectively. This finding is the same as was obtained in the main analysis of the overall Context by Ambiguity interaction above. However, the null hypothesis also could not be rejected for the No Context condition, $z = -0.521$, a finding not in accord with that of the main Context by Ambiguity analysis. Thus, the general Context by Ambiguity effect did not hold for all conditions for all sentences, and the significant triple interaction cannot be completely blamed on the facts of different sentences and different monitor targets in different Groups. Unfortunately, but perhaps not astonishingly, it appears true that not all contexts or all ambiguities were effective for all subjects. Or, put the other way, some materials were more effective than others. Acceptance of this fact seems unavoidable given the triple interaction. The overall Context by Ambiguity interaction (which tests across all materials and all subjects), however, is significant and, as discussed above, is a meaningful and interpretable effect.

---

[2] This analysis was non-orthogonal to the initial analysis, because the data tested earlier were reorganized for this analysis.

Finally, it has been suggested by Foss and Jenkins (1973) that the processing of ambiguities in sentences might differ between subjects who were aware of the presence of ambiguities in the materials and those who were not. In fact, Foss and Jenkins found no difference in the responses of these two groups, but it seems desirable to examine data on this variable in this study, also. (It might be, for example, that only those subjects who were aware of ambiguities would access more than one meaning.) The post-test revealed that 27 subjects reported being aware (at some time during the experiment) that some of the sentences contained ambiguities. The remaining 63 subjects reported that they were not aware of the presence of ambiguities at any time throughout the experiment. Table 2 and Figure 1, respectively, list and display the mean reaction times for these two groups of subjects. While the subjects who were aware of the ambiguities appear to have responded slightly faster in all conditions, the relationships among the conditions of the Context by Ambiguity interaction are remarkably similar for the two groups.

A. Subjects who noticed presence of ambiguities in materials (N=27)

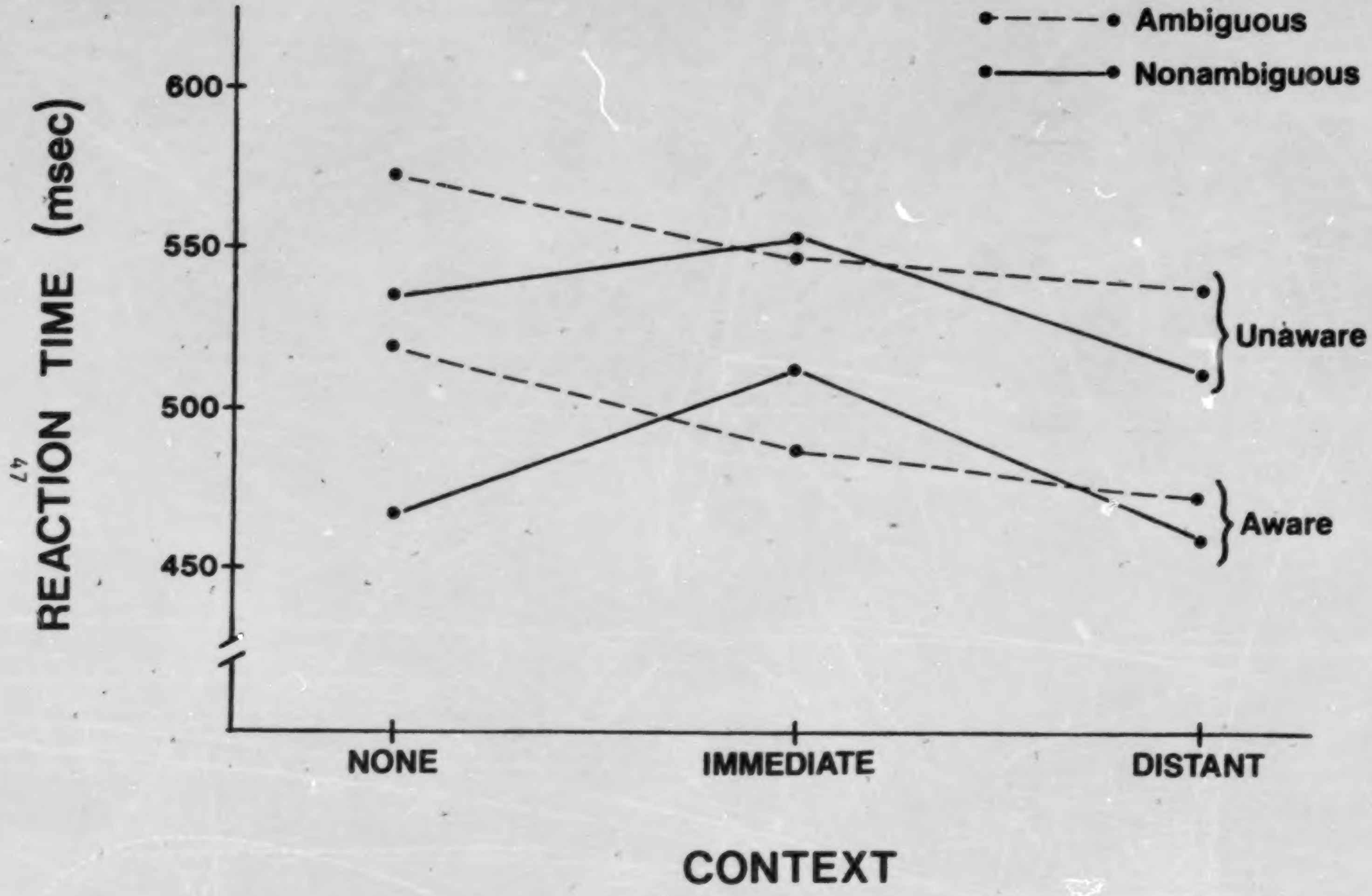
	CONTEXT		
	None	Immediate	Distant
Ambiguous words	520	488	474
Nonambiguous words	469	514	460

B. Subjects who did not notice presence of ambiguities in materials (N=63)

	CONTEXT		
	None	Immediate	Distant
Ambiguous words	573	547	538
Nonambiguous words	536	553	511

Table 2. Mean reaction times (in milliseconds) to the six Ambiguity by Context conditions for (a) subjects who noticed the presence of ambiguities in the materials and (b) subjects who did not notice ambiguities in the materials.

Figure 1. Mean reaction times (milliseconds) to monitor in sentences containing Ambiguous and control(Nonambiguous) words as a function of Context, for subjects who were either Aware or Unaware of the presence of ambiguities in the materials.


Ambiguous
Nonambiguous
REACTION TIME (msec)
47
600
550
500
450
Unaware
Aware
NONE
IMMEDIATE
DISTANT
CONTEXT

An unequal-N analysis of variance[3] was performed, post hoc, comparing subjects who were aware with those who were unaware of the presence of ambiguities as a between-subjects variable. The data were collapsed across the Groups variable of the previous analysis due to the very small numbers of subjects in some groups following the split. Thus, the main variables were Context, Ambiguity, and Awareness. The relevant statistical effects agree with observational judgments made from Figure 1 and Table 2: Awareness was a significant effect, $F(1,88) = 8.906$, $p < .004$; the Context by Ambiguity interaction was significant, $F(2,176) = 3.488$, $p < .03$; and the Awareness by Context by Ambiguity interaction was *not* significant $F(2,176) = 0.283$, $p > .05$. Thus, while those subjects who noticed ambiguities responded 56 milliseconds faster, on the average, than did those who did not notice ambiguities, the basic relations of the significant Context by Ambiguity interaction did not differ for the two groups (as seen in the non-significant triple interaction). This raises the question of why those subjects who did recognized ambiguities responded faster overall. It seems likely that those who noticed ambiguities were the more 'motivated' subjects--ones who tended to respond faster and notice more because they made some effort to do so. What is of particular importance is that the Prior Effect apparently holds up regardless of whether or not ambiguities are noticed. This suggests that we do not have conscious or voluntary control over the processes which are being measured in this experiment.

---

[3] This was based on the analysis described in Winer (1972) where the within group cell mean is substituted for the missing data points. The sums of squares are corrected for inflated N's by multiplying the ratio: $\frac{\text{group harmonic mean}}{\text{pseudo group N}}$. Degrees of freedom are reduced by the number of data points added.

CHAPTER IV

DISCUSSION

I.

The results of this experiment closely bear out predictions derived from the Prior Effect Hypothesis. That is, in the absence of any relevant, biasing context, a lexical ambiguity increases the momentary processing load measured during sentence comprehension. However, a relevant, biasing context occurring prior to the ambiguity reduces this effect. On this hypothesis, the variable reflected by the phoneme monitor task in this particular experiment is the amount of information about the lexical ambiguity which is accessed and made available to the sentence processing mechanism. Under such an interpretation, it can be argued that a relevant, prior context places selective limits (constraints) on what information concerning the lexical ambiguity is accessed.

As previous research has failed to find evidence in accord with a Prior Effect Hypothesis, there is a major question to be answered: what aspects of this experiment distinguish it from previous ones such that the differences in results can be justified? The nature of the context-ambiguity relationships which were used in this experiment appears to be of prime importance in this regard. The contexts were chosen by a rather strict criterion; whenever possible, the strongest a priori and a posteriori relationships were used. Thus, in general, the contexts were closely related (semantically or associatively) to only one reading of the ambiguity, and, in most cases, they eliminated the other potential reading(s) from consideration on the basis of conflicting semantic relationships. Previous research in this area has commonly used contexts which held weak a posteriori (and, often weak a priori) relationships to the ambiguities.

Given the differences between the context-ambiguity relationships used in this and in previous studies, and the related differences in the pattern of results, it seems reasonable to conclude that only certain of the possible kinds of relationships which can obtain between words are sufficient to result in a Prior Effect of context. Perhaps this relationship can best be thought of in the frameword of a strength model. The 'stronger' the context-ambiguity relationship, the higher the likelihood of a Prior Effect. Thus, strong a priori and/or a posteriori relationships are more often sufficient to produce the effect than are weak relationships.[1]

With this model in mind, it appears that we can reject the strongest version of the Prior Effect Hypothesis. That is, while we now have evidence that a Prior Effect can occur, the strongest version of such an hypothesis would claim that any biasing context helps reduce the amount of information about a lexical item which is accessed during sentence processing. Indeed, the failure of previous studies to find a significant Prior Effect is evidence against such a view. Thus, the type of constraints on a Prior Effect Hypothesis implied by a strength model are necessary. The argument for constraints of the sort provided by a strength model also seems to find support on intuitive grounds. Without such a constraint, for example, any word with even the slightest relation to the inappropriate reading of an ambiguity would cause that reading to become activated and the other (perhaps more appropriate) reading(s) to be bypassed. Clearly, one important task is to specify more exactly, the nature of the context-ambiguity relationships sufficient to produce this Prior Effect. The problem will be considered further below.

---

[1] Actually, as has been noted, it is only the strength of the a priori relationship that can be responsible for a Prior Effect of Context. As defined, the a posteriori context-ambiguity relationship is one which allows choice of one of the readings only after all readings are known, generally by eliminating the inappropriate readings on the basis of conflicting or inappropriate relationships. Thus it can only be the strong a priori associative or semantic relationships which result in a Prior Effect.

The question of whether or not the number of words of context is a relevant variable for the occurrence of a Prior Effect of context is related to this consideration of a strength-likelihood model. The Cutler and Foss (1974) experiment, employing sentences containing either one, two or no biasing adjectives prior to an ambiguous noun, studied this variable in a phoneme monitoring paradigm. As already described, interaction between the ambiguity and context in this experiment was not significant, and Cutler and Foss concluded that no prior effect of context was obtained.[2] However, their general results (at least those for ambiguities occurring in the object position) were that the more context words used, the closer the data approached a form which would support a Prior Effect Hypothesis. It seems unlikely, however, that any absolute number of words of context will be shown to provide an all-or-none criterion for this effect (i.e., two words do not produce the effect, but three words do). It does seem likely that the more words of relevant context present, the more likely the conditions necessary for occurrence of the effect will be satisfied.

---

[2] In fact Cutler and Foss (1974) may actually have data which could support a Prior Effect Hypothesis. Their major analysis compared across all context conditions and was apparently designed to determine whether there was an effect of 'amount of context'. To support existence of a Prior Effect, however, all that is needed is to show that a significant ambiguity-control difference disappears under any one of the two context conditions. There appears to be sufficient degrees of freedom to perform these comparisons, but such analyses were not reported by Cutler and Foss. (Of course, such analyses are a matter of choice, and may not have been of major interest to the authors.) I am indebted to D. Hakes for pointing out this possibility to me, although he is innocent of the form in which I have made the argument.

While the materials of the present experiment included more than a single word of context in several of the sentences, the experiment was not designed to study the parameter of 'absolute number of words of context'. However, as many of the sentences in this experiment did contain only a single relevant word of context (See Appendix B), and these sentences did (individually) produce the facilitation effect, there would seem to be corroborative evidence for the hypothesis that it is the nature and degree of relationship between context and ambiguity, rather than any absolute number of related words, that is the important factor. It would seem that, by varying the context-ambiguity relationships through systematic addition of context words with differing, definable relationships, one might successfully isolate the conditions sufficient for occurrence of the Prior Effect. This, however, is easier said than done.

The variable of distance-in-time (or, number-of-syllables) between context and ambiguity also deserves consideration, particularly as previous experiments have not spoken to the issue. The experimental questions related to this variable can be stated simply: will a context occurring immediately in front of the ambiguity be sufficiently processed so as to enable it to produce a Prior Effect on that ambiguity; and, will a context occurring far in advance of the ambiguity have any effect at all, or will some selective mechanism ignore it after a period of time has passed? In short, are there distance-time constraints on the Prior Effect?

There is another, somewhat related, issue concerned with the effect of various hypothetical structural processing mechanisms upon the obtained Prior Effect. For example, based on the supposition that clause boundaries are major areas of sentential processing, it was suggested earlier that lexical items may not be sufficiently processed until the end of a clause to affect processing of later items. If this were true, the context would have to be in the previous clause to be effective.

Variables related to both of these issues were incorporated in the present experiment, but in a manner such that distance was confounded with the presence or absence of a sentence boundary intervening between the context and the ambiguity.[3] In this design, the Immediate Context words were within the same clause as the ambiguity, and occurred from one to three syllables prior to it. The Distant Context words occurred anywhere from 10 to 40 syllables prior to the ambiguity and were in a preceding sentence (see Appendix I for Tables of these distances). If only one or the other of these context conditions had produced results supporting a Prior Effect Hypothesis, interpretation of the source of that effect would necessarily have been compromised and quite difficult. However, as both the Immediate Context and the Distant Context yielded essentially the same effect, discussion is somewhat easier.

To take the second issue first, it seems that the variable of 'clause boundary' may be fairly irrelevant to occurrence of the Prior Effect. The present data indicate that the Prior Effect can occur without the context necessarily being either in the same or in a preceding clause. If, then, there is any particular process of sentential integration or storage occurring within clauses or at clause boundaries, the effects of a prior context upon access of other items is apparently not affected by these. One possibility in accord with this conclusion is that information about a lexical item is accessed immediately when the word is first heard and that such processing does not wait for a clause boundary. This, of course, would not preclude the possibility that important sentential integration procedures might still occur at these boundaries.

---

[3] It is useful in this regard to note that it is extremely difficult to devise long sentences without the use of clause boundaries.

Regarding the first issue raised above, the present data suggest, (although not unambiguously) that simple distance does not differentially alter the processing effects of a prior context on an ambiguity (at least over a mean range from 1 to 18 syllables). Again, while it seems unlikely that prior context effects can be extended over an indefinitely long period of time, it would be interesting to determine if some contexts are effective over very long periods, and, if so, to discover the effective variable in such cases.[4]

Finally, it may be instructive to consider the significant main effect which was obtained for context. In the Immediate Context condition the context apparently led to reduced reaction times for sentences containing ambiguities (compared to the ambiguous, No context, condition), but also led to a relative increase in reaction time for the nonambiguous controls (again, compared to the No context condition). Thus while an

---

[4] It seems likely that a variable of more importance than distance or time is whether or not the subject consciously brings the context to bear on the ambiguity. 'Thematic' contexts, for example, often appear to be effective across extended periods of time. It would be interesting to see if these might produce a Prior Effect of context.

Immediate prior context reduced the processing complexity of ambiguities, it may have actually increased overall sentence processing difficulty. That this might be the case is not surprising when one considers that the Immediate context condition added an extra word (or words) just prior to the point where processing load (complexity) was measured. It seems reasonable that such an addition might either increase or decrease processing complexity at both structural and lexical processing levels, depending on how it is used in interactive processing. That is, while addition of an extra word might have made structural processing more difficult in general, it may have made processing the ambiguity enough easier to offset this effect as reflected in the phoneme monitor task. The Immediate context, nonambiguous control condition does not benefit as much by disambiguation as does the ambiguity, and thus shows a relative loss rather than savings in this condition.

By inspection, it can be seen that both the ambiguous and control conditions are relatively faster in the Distant context condition than in the No context condition. In fact, most of the significant main effect for context appears attributable to the Distant context condition (which differed from the other two conditions by a mean reaction time of from 27-29 milliseconds). If one were to apply the same argument to this case as was used in the Immediate context condition, it could be concluded that the addition of a context word(s) may have made structural processing relatively more difficult in the first sentence; but, such an effect is not reflected in the measure taken in the second sentence. The source of the relative savings in reaction time for both the ambiguous and nonambiguous control conditions, then, would be attributable to the effect of context aiding the choice of a meaning for a word. (In one sense, the unambiguous control word contains only one less source of ambiguity than the so called 'unsystematic ambiguties' and, thus, determination of a meaning, or access, of a nonambiguous word may also be aided by a strong context.) Foss and Jenkins (1973) found a significant main effect for context and attributed it to the ability of context to aid sentence integration. While this is certainly a reasonable explanation, it is at least possible to interpret the main effect as primarily a lexical access effect--as was done here.

The results of the present experiment, then, suggest that, at least in some circumstances, a prior context does facilitate the processing of lexical ambiguities during sentence comprehension. Differences between this and previous experiments suggest that the obtained Prior Effect is parameter bound at least in terms of the strength of the context-ambiguity relationship.

II.

Following Foss and Jenkins (1973) this paper has treated the Prior Effect and Post Effect hypotheses, as presently formulated, as the only possible explanations of prior context effects upon the processing of ambiguous words; these hypotheses have been assumed to exhaust the possibilities for such explanations. The data are clearly not consistent with the Post Effect Hypothesis, and thus it must be rejected. Under the current assumptions, then, it appears that the Prior Effect Hypothesis should be accepted. However, there is some question as to whether this alternative truly exhausts the possibilities remaining for explanation of the present data; while the Prior Effect and Post Effect hypotheses were reasonable first approaches to the context-ambiguity problem, it appears that they may gloss over some important distinctions. To anticipate the argument, it appears that a particular, perhaps incorrect, interpretation of what the phoneme monitor task measures may have been responsible for the original formulation of the Prior Effect and Post Effect Hypotheses. Consideration of alternative possibilities concerning exactly what it is that the task measures suggests that the Prior Effect and Post Effect hypotheses do not exhaust all possible explanations of the present data.

Review of Prior Effect and Post Effect Hypotheses

In order to discuss the reasons for questioning unreserved acceptance of the Prior Effect Hypothesis, it is necessary to recapitulate claims made by the Prior Effect and Post Effect Hypotheses. The Prior Effect Hypothesis holds that context aids one reading of an ambiguous word to become accessed(and therefore activated and available for use by other comprehension processes) and it essentially prevents the other readings from being accessed (and therefore from achieving activation or from being made available to comprehension processes). There are undoubtedly a large number of possible mechanisms which could be posited to implement the Prior Effect Hypothesis. What they all must have in common, however, is the ability to insure that readings of the ambiguity which are not related to a strong context are bypassed, not activated or used, and (metaphorically) never reached by whatever processes are involved in accessing information related to a phonological code. (Thus, this mechanism is not one which reaches a decision between two readings, but rather a filtering process which, guided by context, only allows one reading to be accessed.)[6] The mechanism responsible for this context effect is assumed to act at, or prior to, the stage of lexical access. This is the model (of both the locus and the general nature of the effect) which Foss and Jenkins suggest in their discussion of the Prior Decision Model:

> ...to use a rather crude metaphor, it is as though a light [activation] is turned on next to all related words [to the context] ... When an ambiguous lexical item occurs, the search for its meaning is more likely to eventuate in the related interpretation, the one with the momentary higher level of activation. The lexical search is partially directed by lights next to certain items. (p. 579)[7]

---

[6] However, see Footnote 8 for some qualifications of this claim.

[7] A comment is necessary here. Foss and Jenkins (1973) state that their Prior Decision Model comports with the views of word recognition and attention of Morton (1969) and Norman (1968). Given the prior model, as stated above, this may be a slight misrepresentation of one of these views. The model of Norman (see below for more detail) is a model whereby the context has its effect only after the ambiguity is fully accessed or 'activated'; the context does not selectively direct the access of information. Rather, context aids a (possibly) quicker 'decision' or 'selection' of information than would occur without context, by raising the activation level of information associated with one reading of the ambiguity. This is perhaps a minor point, but it is mentioned here to avoid confusion later when the Norman model is incorporated as the basis of an hypothesis other than the Prior Effect Hypothesis. The essential point of conflict here is over whether the context directs and aids activation of one reading (and prevents activation of the other) of an

This model, and the class of mechanisms that might be posited to account for its operation, bears an interesting and at least analogic, resemblance to a model of attention and selection suggested by Treisman (1960). Treisman, following Broadbent's (early) formulations, suggested that a filtering or selection device exists which can attenuate input signal strength on the basis of context and expectancy immediately after extraction of 'crude physical properties' from the signal but before that signal is accessed in memory. By analogy, it might be posited that nonrelevant readings of an ambiguity are attenuated or blocked prior to their access under a Prior Effect Hypothesis.

When the Post Effect Hypothesis is compared to the Prior Effect Hypothesis, it can be seen that the class of mechanisms which might explain the one is different in nature from the class of mechanisms which might explain the other. The Post Effect Hypothesis, as formulated, holds that all readings of an ambiguous word are accessed (and consequently

---

[7](cont.) ambiguity, or whether it merely acts in conjunction with all readings of the ambiguity which are already accessed. In Norman's model, activation is assumed to occur 'automatically' for all input material and only some of these inputs are selected for further processing. Foss and Jenkins, on the other hand, clearly have their Prior Decision mechanism access and activate only one reading of the ambiguity. They propose a comprehension model whereby: "(1) The information associated with each lexical item in the input string is retrieved or activated." and "(2) This information is then transferred to a limited capacity working memory." They then state that: "The Prior Decision model says that ... Only one meaning is activated and transferred to working memory." (p. 577). This seems to imply that the other reading of the ambiguity is never retrieved (accessed) or activated, and that the context helps it to be bypassed. (Note that, if a reading is accessed it is necessary to posit that it becomes activated above its 'chronic' level in memory (or else flagged or moved to be worked on). Without such an occurrence there is no explanation as to how that reading would be distinguished from any other for later use. The discussion in Foss and Jenkins seems to imply that access (retrieval) results automatically in activation and transfer and that, in the case of a Prior Decision model, only one reading is accessed.)

activated and made available for further processing) and a decision is made among these activated readings. This decision presumably occurs in the same way and at the same time in both the presence and absence of a relevant context (i.e., context affects only the outcome of the decision, not the process by which it is made). Under a Prior Effect interpretation, a different process operates in the presence of a strong context than in the absence of such a context. (This assumption is necessary, because, as mentioned, the Prior Effect can only occur--if at all--in the presence of a strong context.) By way of further comparison, in a Post Effect model a decision is made from among the readings of an ambiguous word, all of which were activated. However, in a Prior model only one reading is ever accessed or activated, and the others are never available. These contrasts form the basis for the claim that the mechanisms for the two models differ in kind.[8]

---

[8] A rather strong claim has been made that, not only are the Prior and Post Effect hypotheses different classes of models, but also the mechanisms that might explain the one are different in kind from those that might explain the other. The Prior Effect is seen as a truly prior-to-access filtering model where one reading is selected and the other is never accessed, and the Post Effect model has all readings used in a decision process. This distinction best preserves the sense of Prior and Post that have been used throughout this and other, related, papers. However, because the essence of the Prior/Post claim is whether or not one or several readings of the ambiguity are made available for further processing, it is possible to envision the mechanics of a Prior system which do not differ from those of the Post system. While (as will be discussed) this change seems to this writer to be a violation of the basic intent of the Prior and Post models, it is necessary to point out that such an alternative version of the mechanics of a Prior Effect model can exist; there are far too few data to constrain the form which the mechanism and theory take. To make this point more explicit, consider that the Prior/Post distinction given thus far holds that when the phonological code is accessed in memory the process of accessing information related to the phonological code, or the result of that process, constitutes 'activation'--just in the sense that the level of 'activation' of the item is raised above 'chronic' level. Anything so activated is considered to be 'available' to the comprehension processes. However, it has been suggested by Foss (personal communication) that it is possible that access of information related to the phonological code does not automatically result in activation. Rather, access results in the 'flagging' of relevant information in memory, where flagging does not constitute activation. Then,

Reconsideration of hypotheses in light of the results of this study: Emergence of a new hypothesis

With this recapitulation of the Prior Effect and Post Effect models in mind, it is now useful to reconsider the data obtained in this study. Following Foss and Jenkins (1973) it was assumed that the dependent variable—reaction time to monitor for a phoneme—reflected momentary processing load at the level of access or activation (or some automatic consequence thereof). Thus, when the present results showed that the processing load associated with an ambiguous word was reduced in the presence of a prior context it was concluded that the amount of information (number of readings) accessed or activated for an ambiguity was reduced in the presence of a strong prior context. Such data seemed to clearly support the Prior Effect Hypothesis, the implication being that information for only one, relevant, reading of the ambiguity was accessed when the Prior Effect obtained.

---

[8] (cont.) in this case, it is possible to conceive of a 'decision' process--much like that posited for the Post Effect mechanism--which is used to activate a relevant reading for an ambiguity (on the basis of a strong context) making it available to later processes, and which does not activate or choose the 'other', unrelated readings, thus disallowing them availability to later processing. Thus, the mechanisms for the Prior and Post hypotheses, contrary to the somewhat stronger claims made above, might be the same. (However, the essential differences in the hypotheses themselves, are, of course, not compromised by the possibility.) In general, it is the version of the mechanics of the Prior Effect Hypothesis which was given in the text, above, which appears to have been intended in previous studies, and which will be used in discussion throughout the rest of this paper.

However, suppose that the assumption concerning the locus of the effect reflected by the phoneme monitor task is incorrect. While Foss and Jenkins (1973) and the present author have (at least implicitly) assigned the locus of the effect measured by the phoneme monitor task to the level of lexical access, not much is known concerning the nature of the phoneme monitor task and, thus, the loci of whatever effects it measures are somewhat underdetermined. Further, unless it can be demonstrated that in this experimental situation the phoneme monitor task is only measuring lexical access or activation per se (or some automatic consequence thereof), the possibility will have to be accepted that it might be measuring some effect of lexical (ambiguity) processing which occurs after access (i.e., after activation of all readings of the ambiguity.) As will be discussed below, there is no evidence which points in either direction, and thus, in addition to the Prior Effect interpretation of these results, an alternative hypothesis can be formulated--one which holds that the phoneme monitor task reflects some processing stage that operates after lexical access and activation and which shows both the effects of increased processing load for an ambiguity, and a decrease in this measure when a strong prior context is present. While this alternative hypothesis is at least similar in kind to the Post Effect hypothesis, it differs in that the Post Effect Hypothesis, as formulated, present no basis for claiming that the resolution of an ambiguity might occur in a different way and at a different time in the presence or absence of a strong context. If the phoneme monitor task is measuring after lexical access, then any 'Post' type model must account for the data obtained in this study. Given that we have rejected the Post Effect Hypothesis, as formulated, the new formulation and alternative hypothesis to the Prior Effect Hypothesis will be called the Immediate Decision Hypothesis. To recapitulate, this new hypothesis holds that the effect of a prior context upon an ambiguous word occurs after lexical access--that is, after all readings of the ambiguity are activated--and that the effect differs from that previously specified by the Post Effect Hypothesis in that it results in the resolution of an ambiguous word more quickly in the presence of a strong prior context than in the presence of a weak context or no context at all. Specifically, the Immediate Decision Hypothesis holds that all readings of the ambiguity are accessed and that a decision process, normally occurring after access, 'decides' upon the correct reading from among all the available readings using context to aid this process; the stronger the context, the faster the decision occurs.

To account for the time course of processing under the Immediate Decision Hypothesis, a decision/selection device is posited to exist following lexical access. Some possible characteristics of this device (called a decision buffer) will be described below. Under the Immediate Decision Hypothesis (as opposed to the Prior Effect Hypothesis) the phoneme monitor task is assumed to reflect processing and storage at this decision buffer. Similarly, the locus for the effect of context is at this decision buffer, following lexical access and activation of all readings of the ambiguity. Note that the three hypotheses discussed thus far in the paper hold an interesting relationship. The Prior Effect hypothesis implies that the way in which an ambiguity is resolved in the presence of a strong prior context is different in nature from the way in which the ambiguity is resolved without a biasing context; a strong context selects (or predisposes selection of) one reading, but with no context (or a weak context (or a weak context) a decision is made among all available readings. The Post Effect Hypothesis, on the other hand, holds that ambiguities are resolved by the same process in both the presence or absence of a strong context. It is quite vague as to when this resolution occurs, but the implication is that, regardless of context, it always occurs at the same time in the same way. The Immediate Decision Hypothesis, like the Post Effect hypothesis, holds that the same process is involved in resolving ambiguities with or without strong prior contexts, but claims that a strong context will cause the process to be invoked earlier than will a weak context or no context. Thus, it is in terms of the time of occurrence of the effect in which the Immediate Decision and previously specified Post Effect Hypotheses differ. Quite obviously, the Immediate Decision Hypothesis is merely a particular and more explicit instantiation of a, thus far, vaguely and incompletely specified Post Effect Hypothesis.

The problem now to be faced is one of determining which of these two hypotheses--Immediate Decision or Prior Effect—the data from this experiment truly support. It is an unfortunate but presently unavoidable state of affairs in psycholinguistics in general, and in comprehension research in particular, that little is known of the exact nature of our behavioral measures.[9] Clarification of what the present data mean requires more explicit specification of the mechanism or process which is being reflected by an increase in phoneme monitor reaction time in the presence of an ambiguity. Note that the number and types of processes potentially reflected by a phoneme monitor task encompass a wide range of possibilities. In general, the task most likely reflects some process or processes occurring concurrently with, and at least partially shared with, processes involved in the recognition search for a phoneme. Thus, the phoneme monitoring processes and the lexical access and/or lexical decision processes need only share some mechanism (any one of the whole series of mechanisms utilized for processing words in sentences). With the exception of that shared mechanism they could be essentially separate and different processes. In short, it is nearly impossible to determine, at this time, exactly where and what the phoneme monitor task is measuring. While it might be hoped that past studies using this technique would have shown the parameters which it is capable of measuring to be quite restricted, that is not the case. The task has been found to reflect effects that are strictly lexical in nature (e.g., Foss, 1969) and those of structural (presumably integrative) differences in sentence processing (e.g., Hakes & Foss, 1970). Thus, in the ambiguity experiments the task could be reflecting some aspect of accessing lexical information, or it could equally well be reflecting some later stage of processing. The fact that the measured effect occurs quite closely following occurrence of the ambiguous word itself requires, of course, that any such later processing stage be one that is reached quickly.

---

[9] See Foss, in preparation, for further comments on this problem.

Certainly both the Prior and Immediate Decision Hypotheses are equally valid explanations of the present data. There is, however, some (albeit weak) motivation for questioning whether the class of models suggested by the Prior Effect Hypothesis are reasonable.

First, recall that the type of context used in this experiment not only held a strong a priori relationship to the ambiguity, but also a strong a posteriori relationship. The a posteriori relationship is that which allows for selection of one reading of an ambiguity after all readings are known, by virtue of conflicting relationships with other readings. The potential importance of this relationship has been ignored in the discussion thus far, but, because previous studies often used contexts which were weak in this property it is possible that it was the a posteriori property of the context-ambiguity relationship that resulted in the obtained context effect. If so, it could only have had its effect after all readings were accessed and activated.

In addition, consider again that while the Prior Effect Hypothesis predicts that the presence of a strong context will eventuate in lexical access and activation of one reading of the ambiguity, it also predicts that the other reading(s) of the ambiguity is (are) bypassed and never accessed. If true, it is surprising that there is not more garden pathing than there seems to be in language processing. Because the Prior context process must operate blindly to eliminate one or more readings without ever knowing what they are, and because the context effects are reasonably long lived (context in this study was, at maximum, 40 syllables prior to the ambiguity), it is amazing that the mechanism is not mislead by some context or other (perhaps intended to realte to another topic in another sentence) more often than seems to be the case.

Finally, there is an argument, made earlier, which supports the Immediate Decision Hypothesis by appeal to a sense of parsimony. It seems to make less intuitive sense to have two separate processes act upon ambiguities--one in the presence of strong prior context and one in its absence--

than for a single process to deal with ambiguities under both conditions.

In all, there is certainly no proof that the context effect obtained in this study is not the Prior Effect it has thus far been assumed to be; context may have its effect prior to or at lexical access, and it may well prevent activation of the unrelated readings of the ambiguity. But the evidence does not compel such an account.

The Immediate Decision Hypothesis (as did the Prior Effect Hypothesis) bears at least analogic resemblance to several current models of attention, selection, and word recognition. These models suggest some interesting mechanisms which might operate under an Immediate Decision Hypothesis. Norman (1968) for example, has a model of attention/selection in which an input stimulus undergoes an initial analysis to extract sensory (physical) features of the signal; these, in turn, activate information at a memory address determined by the form of the features themselves. Selection of some part of the activated information for further attention occurs through the use of a device which selects the most highly activated information for further analysis. The level of activation of information can be raised by a pertinance mechanism (context). Thus, this model allows for selection only after all input information is activated, context providing an additional source of activation that can have its effects on the information to which it relates.

The model of Norman (1968), as well as that of Morton (1969), contains several features which have interesting possibilities for application to the decision buffer which has been posited for the Immediate Decision Model. Some of the more relevant of these features include: 1) a suggestion that, whatever the source of the information, there is a common code or format that all information shares, thus enabling different pieces of information to be related to each other (this appears to be a logical necessity for memory in general, and not one just related to the mechanism under discussion); 2) all input information activates its representation in memory (even unattended information); 3) when pertinence (context) excites a particular representation in storage, it must activate it for all items associated with it through the 'interconnected structures or storage'; and 4) entries in memory have some basic level of activation (perhaps related to frequency of occurrence[10]) and context acts to raise the level of activation of information to which it relates. Morton further suggests that the first piece of information which passes some criterion level (threshold) is made available for further processing and the rest is not.

Incorporating these features into a larger and more specific scheme, it is possible to lay out an interesting picture of how the decision buffer posited for the Immediate Decision Hypothesis might operate. The relevant defining data are so few that any description of the mechanism at this point could only be one of a very large number of possibilities. However, making the model more explicit can only make its verification or refutation easier; thus, something may be learned from this attempt.

---

[10] In Morton's (1969) model, frequency adjusts the threshold which must be attained for the item to become available for further processing rather than actually changing the level of activation. (High frequency items have their threshold moved downwards, low frequency items have it moved higher.)

Consider the processing of an ambiguous word. First, access of the phonological code causes all (let us say two) readings of the ambiguity to become activated. Access is assumed to raise the level of activation by a constant amount 'X' above the chronic level of activation of the item in memory. The relative difference between the chronic levels of the two readings is a function of the frequency or bias associated with each reading. Then, all readings are made available (transferred) to a limited capacity decision buffer, each at its current level of activation. It is at this stage that context interacts with these readings. Readings related to the context will receive an increment in activation, with the amount of increment depending upon the strength of the relationship. Information for each reading is held in the buffer until one reading is passed on, or made available, to further processing. As implied by this last claim, there is a threshold on the output of the buffer. The first reading of the ambiguity to attain a level of activation exceeding the threshold level in the buffer is passed on. Once one reading passes threshold, the buffer is cleared (and the next available word enters). (In clearing the buffer, the 'activated' information is returned to memory.) Further, to capture the fact that even in the absence of contextual activation decisions are eventually made in resolution of the ambiguity, it is posited that there is a time dependent change in the threshold level in the buffer. The threshold is set high initially with a downward criterion employed to lower the threshold as time passes. Thus, the stronger the relevant context, the more quickly one reading is gated onward. This criterion shift would allow for the Cairns (in preparation) data which indicate that ambiguities are resolved fairly quickly (but not immediately) even with no biasing context, as well as for the results of the present study which found an (almost) immediate effect of a strong context upon resolution of an ambiguity. Further, the phoneme monitor task is seen to reflect processing at this decision buffer. There are a number of ways in which this might occur. For example, it could be that the monitor uses the mechanism (or some of the processes of that mechanism) of the buffer. When a large amount of information is present in the buffer, more of the precessing mechanism is utilized than when less information is stored in the buffer. Thus, the phoneme monitor task, needing to use some part of this same mechanism, is sensitive to the amount of information in the buffer; an ambiguous word will cause a longer delay in a phoneme monitor task than will an unambiguous word.

It can be seen that this posited mechanism accounts for all the known and related data--those of Foss (1970), Foss and Jenkins (1973), Cairns (in preparation) and the present experiment. Even so, it should be reiterated that this is just one of a large number of possible models for this effect.

Following this rather lengthy discussion of the Immediate Decision model, it needs to be again stated that the data do not unequivocally support either this hypothesis or the Prior Effect Hypothesis. While one may seem more reasonable than the other, an empirical test between these is still necessary. It seems that the crucial information which must be focused on in such a test is whether or not the 'other' readings of the ambiguity are activated in the presence of a strong prior, biasing context; apparently a measure other than that presently used is called for in this situation.

CHAPTER V

SUMMARY, CONCLUSIONS, FUTURE RESEARCH

The present research investigated two basic hypotheses concerning the nature of the effect that a prior lexical context has upon the interpretation of a lexical ambiguity during sentence comprehension. The hypotheses were based on the results of various measures of comprehension which had found evidence of greater processing complexity in the presence of unbiased, lexical ambiguities than in the presence of nonambiguous words; this greater complexity was taken as evidence that more than one reading of an ambiguity is made available to comprehension processes under unbiased conditions.

According to one of the two hypotheses under investigation (the Prior Effect Hypothesis) a prior context has its effect upon an ambiguity before the ambiguity is processed; it predisposes the processing mechanism toward the choice of the related reading of the ambiguity, and effectively bypasses the unrelated reading(s). The Post Effect Hypothesis, on the other hand, would have the influence of a prior context occur only after all information about the several readings of the ambiguity were accessed in memory.

The studies by Foss and his associates, using one of the more sensitive measures of on-line processing complexity (the phoneme monitoring task), found that the relative processing complexity associated with unbiased ambiguities did not disappear in the presence of a prior context. It was assumed that the phoneme monitor task reflected processing load at lexical access (or, alternatively, at a working memory to which the lexical item was immediately transferred following access). Thus, as did other studies using less sensitive measures, Foss and his associates' work appeared to support the Post Effect Hypothesis.

In spite of these results, the present research was motivated by an interest in whether the Prior Effect Hypothesis might not still find support, even if only in limited conditions, and an interest in the effect of certain sentence processing variables upon context-ambiguity effects in general. The basis for these interests lay in the belief that important variables had been overlooked, and some necessary controls omitted, in previous experimental studies. Of particular importance in this regard was consideration of the strength of the context-ambiguity relationship, in two specifically defined areas, the distance of the prior context from the ambiguity, and the relative effect of a context occurring either in the same sentence as the ambiguity or in a sentence previous to it.

The results of this study support the Prior Effect Hypothesis; the significant difference in reaction time between the ambiguity and context conditions in the absence of a biasing context disappeared in the presence of a biasing context. This result held for contexts which were in the same clause as the ambiguity and, on the average, 1.3 syllables in front of it as well as for contexts which were in the sentence previous to the one containing the ambiguity and an average of 18 syllables prior to it. Thus, it appeared that the lack of support for the Prior Effect Hypothesis in previous experiments was due, at least primarily, to their failure to use appropriate context-ambiguity relationships. On this basis, a somewhat constrained model of the Prior Effect Hypothesis was suggested wherein certain strong context-ambiguity relationships would have to hold before context could have this effect.

It was pointed out, however, that this particular interpretation of data relative to the Prior Effect and Post Effect Hypotheses presupposed knowledge of where the phoneme monitor task was tapping in on the comprehension process. As already mentioned, this was assumed to be at lexical access or at a 'working memory' to which activated information was transferred immediately following access (the implication being that everything activated was transferred). An alternative view of where the obtained effect might be occurring was presented in which it was suggested that the monitor task could well be measuring processing load at some point following lexical access or activation. In such a case, the reduction in processing complexity found in the presence of a biasing context could be the result of a decision process occurring after access of all information associated with the ambiguity, a decision which eventuates in only one reading of the ambiguity being chosen for further processing and one which occurs more quickly in the presence of 'strongly' related contexts than weakly related ones. This hypothesis was labeled the Immediate Decision Hypothesis. It differs from the Post Effect Hypothesis primarily in terms of the timing of its effects.

Under the Immediate Decision Hypothesis interpretation, a decision buffer with various processing characteristics was posited to exist immediately following lexical access; resolution of the ambiguity was suggested to occur at this buffer. Relatedly, it was suggested that the phoneme monitor task may have been measuring processing which occurred at this point.

In all, this study has shown that prior lexical contexts can reduce the processing complexity of lexical ambiguities, as measured by the phoneme monitor task, during sentence processing—a conclusion rejected in previous research. This effect is apparently related to the strength of either the a priori or a posteriori relationship of the context to the ambiguity, and appears to be relatively free of simple effects of context-ambiguity distance or intermediating clause boundaries. What is left uncertain is the exact locus of the effect. Thus, perhaps the most important problem in need of resolution is whether or not the effects obtained in this study can be attributed to processes modeled in the Prior Effect or Immediate Decision Hypotheses. In one sense this is a question of where the phoneme monitor task is measuring whatever it measures.

It can, however, also be attacked from another approach. The Prior Effect Hypothesis predicts that only the 'related' reading of the ambiguity will be accessed and activated and other readings will not. On the other hand, the Immediate Decision Hypothesis predicts that both readings are accessed and activated even in the face of a strong context. Thus, some measure of whether or not the unrelated reading(s) is ever activated and available for further processing in the presence of a strong prior context would indicate which hypothesis of the effects of prior context is correct.

Regardless of which model of the results is correct, description of the exact processes involved requires determination of the characteristics of the context-ambiguity relationship which were sufficient and/or responsible for the context effect found in this experiment. For one thing, the a priori and a posteriori context-ambiguity relationships need to be studied independently. Considered more generally, specific information concerning what type of information can be used as a context is needed.

For example, can structural (syntactic) information act as a context. Structural information (such as class occurrence restrictions) is at least part of what must be retreived when information about a word is accessed. It seems possible that this information could be the basis for the type of effect predicted for context under either an Immediate Decision Model or a Prior Effect Model; if the next word in a sentence can only be part of speech X, then this information could be used to activate (select) a reading which has the property X. There are various models of parsing routines which have properties of both predictability (expectation) and of delayed-decision-making built into their routines. For example, Kimball (1973) suggests that occurrence of a grammatical function word signals the necessity for construction of a new structural node. In this case, structure is being predicted (although the exact form class for the next word is not). This same model, however, claims to be, in general, a look-ahead model—one which assumes that a listener may wait a given number of symbols subsequent to a particular terminal symbol before assigning it a particular structural description. There exist models which exemplify the predictive and delayed-decision characteristics more strongly, but for the time-being it is sufficient to note that the possibility exists that structural requirements could act as a prior context in the decision/selection process for lexical items; it is now a matter for empirical investigation.

Finally, there is a consideration which was hinted at earlier in this paper. It is the question, given the possibility that the Immediate Decision Hypothesis is correct, of whether or not the lexical decision processes could apply during access of nonambiguous lexical items as well as ambiguous ones. As was mentioned earlier, there is apparently much more information related to a lexical item than is (consciously) retrieved during sentence comprehension. The question, then, is what is the basis for selection of some subset of this information? It appears that the same type of decision must be made for nonambiguous words as is made for ambiguous words; only some fraction of the information actually related to the lexical item is usable and fulfills the needs of any one situation. Recall that in the present study the presence of a prior context resulted in reaction times which were relatively faster even for nonambiguous control conditions (this only held for the Distant context condition, which, as was explained, was the only condition unconfounded by presence of an additional word in the sentence where the phoneme monitor task was measuring complexity. Thus, it could be that the context was involved in a decision process which occurs for all words--ambiguous or nonambiguous.

It is often stated that one sign of a succesful research project is when it raises more questions than it answers. Under this criterion, at least, this study has had some measure of success; the unanswered questions abound.

APPENDIX A

FREQUENCY (THORNDIKE-LORGE COUNT) AND LENGTH (IN SYLLABLES) OF AMBIGUOUS WORDS AND UNAMBIGUOUS CONTROL WORDS USED IN EXPERIMENTAL SENTENCE (TI)

SENTENCE #	AMBIGUITY			UNAMBIGUOUS CONTROL		
	word	freq.	length	word	freq.	length
1	records	AA	2	reports	AA	2
2	stand	AA	1	belief	AA	2
3	glasses	AA	2	containers	AA	3
4	pipe	A	1	cigarette	22	3
5	pot	A	1	pan	47	1
6	race	AA	1	track	A	1
7	tanks	19	1	trucks	23	1
8	club	AA	1	group	AA	1
9	plays	AA	1	operas	36	2
10	toast	21	1	biscuit	14	2
11	plants	AA	1	trees	AA	1
12	duties	AA	2	job	AA	1
13	ball	AA	1	dance	AA	1
14	conductor	16	3	musician	18	3
15	class's	AA	2	group's	AA	1
16	tables	AA	2	desks	A	1
17	seals	45	1	bindings	47	2
18	board	AA	1	committee	A	3
19	punch	17	1	liquor	34	2
20	bulbs	14	1	seeds	A	1

21	boxer	3	2	poodle	1	2
22	iron	AA	2	clothes	AA	1
23	interest	AA	2	dividends	A	3
24	stock	AA	1	animals	AA	3
25	panel	11	2	shelves	11	1
26	bow	A	1	arrow	37	2
27	bugs	10	1	insects	10	2
28	drilling	21	2	digging	43	2
29	speaker's	AA	2	teacher's	AA	2
30	note	AA	1	song	AA	1

APPENDIX B

LIST OF EXPERIMENTAL SENTENCES IN SCHEMATIZED FORM

1. It seemed that (bookkeeping) materials in the small store were always
in a mess. The store manager planned to organize the (bookkeeping)
records during his lunch hour.
reports Δ

2. From recent experience, the man had found that few people liked his
(philosophical) work. In hopes of interesting someone in it, he
described his (philosophical) stand carefully to those present.
belief Δ

3. The supper guests were surprised by the presence of the old (tobacco
smoking) equipment in the next room. Everyone gradually became aware
of the unpleasant odor (of tobacco smoke) from the pipe in the
Δ dining room of that old house. cigarette

4. An unexpected shortage (of deep fry, cooking equipment) arose in the
small town. The young man delivered the (deep fry, cooking) pot to
his best customer, and the others had to wait. pan
Δ

5. The (diesel-powered, military) equipment had been made to take abuse,
but it was now very old. The authorities inspected the (diesel-
powered military) tanks carefully to determine if they needed repairs.
trunks Δ

6. Most interested friends felt that his decision (to join the elite fraternal
society) was unfortunate. But, the young man had thought carefully
before deciding which (elite, fraternal) club to pick as the one best
suited for him. group Δ

7. Even though Tom had long been associated with the (ballet and theatrical)
school, he still was not immune to criticism. His critics contended that
his choice of (ballets and theatrical) plays generally had been terrible,
operas Δ
as had been the case with all his other decisions.

8. Most of the guests had looked forward to the (homemade jam at the) meal
with enthusiasm. The lack of any (homemade jam and) toast during the
entire meal was surprising to them. biscuits Δ

9. That town was known locally as a good place for (flower and shrub)
businesses to begin. Year around good weather kept the cost for develop-
ment of (flowers, shrubs, and) plants quite low in that area.
trees Δ

10. Information had been only recently handed down to the local officials
about (responsibilities under) the new regulations. The officials felt
that the (responsibilities involved in the) new duties would prohibit
enterprise and stifle free development. job Δ

11. The manager (of the Fred Astaire Studio) spent a long time in presenting
the reasons for the change he suggested. Even so, most persons were
upset over the problems which would be incurred in having a (Fred
Astaire Studio) ball of the cost which he had outlined.
dance △

12. Several people always arrived early to watch the (renouned soloist and
his) unusual preparations. Everyone then watched as the (renouned
soloist and excellent) conductor bent to his tasks.
musician △

13. Rumors concerning complaints among the (high school) members had
spread, but the causes varied. Among the ones the ones the young
man heard were unfavorable comments concerning his (high school)
class's donation to the socialist movement.
group's △

14. The man was worried that the quality of his work on the (antique
chairs and other)items might not be acceptable. Several men had
come by at his request to inspect the (antique chairs and) tables
prior to completion of the work. desks
△

15. In the last few days there had been some difficulties with the
(customs inspection stickers on) containers during shipment. The
officials checked the (customs inspection stickers and) seals
prior to making the report to their superiors. bindings
△

16. Early in the week John received a telephone call from his friend
requesting his advice (concerning the new executive director
position). He responded with this suggestion for the (new
executive direction and for the) board very quickly, but
committee △
retained some uncertainty over his choice.

17. In the early evening, the retired gentleman had joined in the (wine
sampling) activities. The old man felt the effects of the (wine and
the) punch during the rest of the evening.
liquor △

18. The (South Austin Kennel Club's) entry had long been heralded as a
potential champion. Everyone eagerly watched the (South Austin
Kennel Club's) boxer during the early trials.
poodle △

19. Upon entering the shop, Fred spotted the (cleaner's steam-pressing)
items which he had to pick up in the corner. Fred felt certain that
the (clearner's steam-pressing) iron would be difficult to move.
ed clothes △

20. All day long the situation (on rates for stocks and bonds) had been
in a state of flux. The last announcement had lowered (rates for stock
and bond) interest by a greater amount than was expected.
dividends △

21. John's partner objected to his making the purchase (of the cattle)
because of the high cost. But, John finally bought the (cattle and
other) stock despite all the objections.
animals Δ

22. Paul wasn't sure that anything could change the arrangement (of the
wooden walls). Considerable pressure had to be applied before the
(wooden wall) panel gave way in the desired direction.
shelves Δ

23. Bill suggested that some new (wooden, archer's) equipment was needed.
He thought that her (wooden archer's) bow didn't seem appropriate
for the situation. arrow Δ

24. Rumor had it that for years the government building had been plagued
with (spiders, roaches, and other) problems. The man was not surprised
when he found several (spiders, roaches, and other) bugs in the
corners of his room. insects
Δ

25. The government officials had decided to continue (laying the new pavement) according to old procedures. The men (laying the new pavement)
were drilling in the sun all during the long hot day.
digging Δ

26. It had been rumored that her (melodic alto) work was not up to par.
Therefore, Jack was surprised at the quality of sadness contained
in her final (melodic, alto) note despite having long known her
abilities quite well. song Δ

27. Her sister had left (the flowers, seeds and other) things in the
wrong place. The young lady took the (flowers, seeds and) bulbs
downstairs to store them where they belonged. seeds
Δ

28. The elderly woman finally went to purchase the (formal chinaware
and crystal) replacements that she had needed for a long time. She
was amazed at the variety of (formal chinaware and crystal) glasses
downtown at the main branch of the store. containers
Δ

29. It appeared that a lot of consideration had been given to making
conditions for (both the marathon and sprint) entries as good as
could be expected. Thus, writers were a bit surprised to hear
rumors of (both marathon and sprint) race problems at the Olympic
Games. track Δ

30. The group of experts invited to the meeting agreed that the unusual
(speech and elocutionary) comparison method was quite useful and novel.
They found they enjoyed discussing (speech and elocutionary) aspects
of each speaker's performance using the 'dual comparison' method.
teacher's Δ

APPENDIX C

FILLER SENTENCES WITH MONITOR TARGET OCCURRING EARLY IN THE SECOND SENTENCE

Phoneme Target		
/s/	1.	Home economics was a class traditionally taught by persons with little academic training. Since sex education became a part of home economics, the teachers have had much more technical education.
/m/	2.	Cigarette smokers have long been a bother to non-smokers. However, many workers who gave up cigarettes turned to cigars, which non-smokers find even more obnoxious.
/ch/	3.	A lawyer filed court papers for an unidentified client. The charge against the defendant was suddenly dropped, causing rumor of a fix.
/v/	4.	The life of a Taxi driver might seem to be dull. The variety of life one sees and the money one makes while driving a cab apparently offset the disadvantages for many taxi drivers.
/k/	5.	The weatherman predicted no rain for that week. The week-long canoe trip was postponed due to terrible weather.
/n/	6.	The former prison guard was interviewed at length. He claimed no knowledge of the planned insurrection.
/f/	7.	House appropriations committee members vowed to reduce spending at that session. The funds for education were slashed drastically for the next two years.
/d/	8.	Now that China has been admitted to the U.N., world trade is expected to develop in new directions. Many economists don't feel that the U.S. will gain much from trade with far eastern countries.
/t/	9.	The restaurant owner hated to accept checks. The town was becoming known as a center for hot check passing.
/b/	10.	The workman took pride in his trade. The bootmaker had become famous for comfortable, carefully made products.
/b/	11.	Fred noted that the two people were too preoccupied with each other to notice him. As long as they sat by the fire, he felt all would be well.

/k/ 12. The favorite night spot has just made an exception to a long standing rule and started using female entertainers. Newspapers reported that the ladies undeniably brought class to the night club act.

/k^{w}/ 13. The block was full of young growing children. However, the lady maintained that her twins grew quicker than any of the other children.

/d/ 14. They were warned that she liked to play both sides of the fence. However, both Fred and John thought Doris liked only himself.

/t/ 15. The gym class of overweight persons gathered once a week for work-outs. Even though they all got tired quickly, the instructor insists that they continue running.

/d/ 16. The young researcher taking care of the children cursed all the time. Then he reported that children said dirty words when watching violence on television.

/k/ 17. Solicitors often approach people at offices and again at home. Therefore, businessmen always say they give contributions to needy organizations at the office.

/s/ 18. The group had no choice but to hire non-union workers. They were amazed to find that non-union construction men build structures with as great speed and care as union workers.

/s/ 19. For a long time we didn't understand what an opportunity he had. His friends tell stories of how David finally left his job to go where he felt comfortable.

/d/ 20. Police brutality was claimed in the case. But, the police shake down all the apprehended felons as a matter of course, so the judge ignored the claim.

APPENDIX D

FILLER SENTENCES WITHOUT TARGET PRESENT

Phoneme Target		
/b/	1.	It was often the case that Loretta ate dinner alone. That evening, however, she had hopes that a friend would call to take her out.
/g/	2.	Most of the students worked with enthusiasm. However, two of the boys in the class consistently wasted time playing tic-tac-toe.
/d/	3.	The railroad workers tried to organize a strike. It had been rumored that owners of the corporation were about to let the railroad fold.
/r/	4.	The comic book artist finally give up his line of work. The man found that he had begun to dress like the character he drew.
/k/	5.	Professors have long been notorious for having poor memories. The new department chairman was unhappy to find that his professors recalled his promises all too well.
/m/	6.	The editors of the daily newspaper claimed to be impartial. It surprised no one, however, when their editorials continually supported only one side of the issue.
/k/	7.	Manufacturers of prescription drugs have had few restriction on their prices up to this time. The national meeting of drug makers set up a lobby which would bargain for reduced taxes.
/m/	8.	All too often, the darlings of the jet set seem to turn into disappointed and weary people. Far before their time, they show the signs of age.
/g/	9.	Some of the proceedings from the event went to charity. Most of it, however, was pocketed by the promoters.
/p/	10.	The handsome bellhop escorted the little old lady to her room. The hotel had a reputation for having dazzling service.
/k/	11.	The boxer and his trainer hated morning work-outs. Preparation for the upcoming fight had been rather sloppy.

/d/ 12. Most boys have a tendency to tease girls when they are young. The older man, however, appeared to find teasing as important a part of life now as it was for him in years past.

/k/ 13. The producer and his associates had never staged a successful production. It appeared that they were doomed to be failures in all of their enterprises.

/d/ 14. Spanish style furniture has become very fashionable in the United States. The trend in mass manufacture of furniture is to make cheap Spanish trim.

/g/ 15. The lady didn't know which salesman to listen to. The tire salesman from two different companies has given her totally opposite advice.

/n/ 16. Professional sports used to be very risky investments. The broker advised her client that investing in professional basketball was quite safe these days.

/m/ 17. The salesman quickly became certain that his pitch was perfect for the job. He had talked 10 people in that area into buying things they didn't need.

/b/ 18. The writer had not produced anything for nearly six months. The royalties from his last play had made life easy.

/s/ 19. The president of the university was caught unprepared for the news. In his absence the board of regents had voted to change major university policy.

/t/ 20. It is often supposed that psychiatrists are sensitive, concerned persons. A magazine article showed many psychiatrists were interested only in making money.

APPENDIX E --- SAMPLE RESPONSE SHEET

Subject's name ____________________ SS# ____________________ Exp. Group# __________

	VR	MR	N	MU	VU
1.	____	____	____	____	____
2.	____	____	____	____	____
3.	____	____	____	____	____
4.	____	____	____	____	____
5.	____	____	____	____	____
6.	____	____	____	____	____
7.	____	____	____	____	____
8.	____	____	____	____	____
9.	____	____	____	____	____
10.	____	____	____	____	____
11.	____	____	____	____	____
12.	____	____	____	____	____
13.	____	____	____	____	____
14.	____	____	____	____	____
15.	____	____	____	____	____
16.	____	____	____	____	____
17.	____	____	____	____	____
18.	____	____	____	____	____
19.	____	____	____	____	____
20.	____	____	____	____	____
21.	____	____	____	____	____
22.	____	____	____	____	____
23.	____	____	____	____	____
24.	____	____	____	____	____
25.	____	____	____	____	____
26.	____	____	____	____	____
27.	____	____	____	____	____
28.	____	____	____	____	____
29.	____	____	____	____	____
30.	____	____	____	____	____
31.	____	____	____	____	____
32.	____	____	____	____	____
33.	____	____	____	____	____
34.	____	____	____	____	____
35.	____	____	____	____	____
36.	____	____	____	____	____
37.	____	____	____	____	____
38	____	____	____	____	____
39.	____	____	____	____	____
40.	____	____	____	____	____
41.	____	____	____	____	____
42.	____	____	____	____	____
43.	____	____	____	____	____
44.	____	____	____	____	____
45.	____	____	____	____	____
46.	____	____	____	____	____
47.	____	____	____	____	____
48.	____	____	____	____	____
49.	____	____	____	____	____
50.	____	____	____	____	____
51.	____	____	____	____	____
52.	____	____	____	____	____
53.	____	____	____	____	____
54.	____	____	____	____	____
55.	____	____	____	____	____
56.	____	____	____	____	____
57.	____	____	____	____	____
58.	____	____	____	____	____
59.	____	____	____	____	____
60.	____	____	____	____	____
61.	____	____	____	____	____
62.	____	____	____	____	____
63.	____	____	____	____	____
64.	____	____	____	____	____
65.	____	____	____	____	____
66.	____	____	____	____	____
67.	____	____	____	____	____
68.	____	____	____	____	____
69.	____	____	____	____	____
70.	____	____	____	____	____

EXAMPLES

1.	____	____	____	____	____
2.	____	____	____	____	____
3.	____	____	____	____	____

QUESTIONS

1.	____	____	____	____	____
2.	____	____	____	____	____

*KEY: VR=very related; MR=mildly related; N=neutral; MU=mildly unrelated; VU=very unrelated

APPENDIX F

INSTRUCTIONS GIVEN TO Ss

This experiment is investigating some aspects of language comprehension. It is not a test of intelligence or personality, and we will make no attempts to mislead you in any way. What you will be asked to do is to listen to some sentences and to perform two tasks while doing that. Specifically, you will hear a large number of sentence-pairs; that is, two somewhat related sentences, such as this example:

> 'The young golfer practiced two hours each day. His coach told him that he would be good enough for the pro ranks in a year.'

Before you hear each sentence-pair, you will hear a phoneme, that is, a letter sound, such as /b/ in 'boy'. One of your tasks is to listen for that phoneme beginning some word in the sentence pair. As soon as you hear that phoneme beginning some word, press the button in front of you as quickly as you can. Use either button and use your preferred hand, but be consistent - use the same hand and button throughout the experiment. Please rest your index finger lightly on the button at all times during the experiment. The target phoneme will only occur in the second sentence of each sentence pair. It will never begin any word in the first sentence, so do not be concerned with listening for it there. The phoneme can occur at the beginning of any word anywhere in the second sentence, however. Note that if the target is /b/ as in 'boy', you would respond if you hear the word 'batter', but not if you hear the word 'rabbit', as the word 'rabbit' does not begin with the phoneme /b/. In some sentences, the specified target phoneme will not occur at all. This is to keep you honest. Obviously, you are not to respond on these trials.

Let us try an example of phoneme monitoring, with the example you heard a minute ago. Ready? Listen for the /k/ sound as in 'cat'. Ready?

(repeat the example above)

You should have pressed

Finally, you must be certain to listen carefully and understand the sentences as you hear them. Your second task is to mark, on the sheets of paper we have provided for you, how related you feel the two sentences of each sentence-pair are. The scale is from 1 to 5, and stands for the strong related to weakly related continuum. All of the sentences have something in common, but some will relate more closely than others. We will not specify any particular dimension of relatedness on which to judge the sentences. There is no absolute or correct dimension, so any you choose to use is acceptable. Do not be concerned if you have trouble finding the relative scaling of pairs at first. Eventually you will probably have a few dimensions on which you can relate relatedness. Please place a slash across one of the five points on the scale for each sentence pair indicating the relatedness of the members of the pair. Please use the hand not resting on the response button. There will be only a short interval in which to make decisions following each sentence-pair, so be certain to listen to the sentences carefully and understand them both, so that you can make your decision rapidly. Once you have made a decision do not change it.

Thus, to summarize, you will hear a number of sentence-pairs. Listen carefully and understand the sentences. A phoneme will be specified before each sentence-pair. You are to press the button in front of you as soon as you hear that phoneme in the second sentence of the pair of sentences. Listen carefully, as it is often difficult to spot the phoneme. Also, please press the button as quickly as you can, as we are very interested in reaction time. Following that, you are to place a slash on the sheets provided to you indicating how related you feel the two sentences are. Are there any questions? Let's try some examples to get a feeling for what the task is like. Please press your buttons and mark the slashes just as though this were the real experiment. Ready? Listen for the /k/ sound as in 'cat'. Ready?

> 'The electrical appliances on the store's shelves were beautiful to look at. The manufacturer had been taken to court by thousands of people for malfunction of their appliances.'

Listen for the /g/ sound as in 'gun'. Ready?

> 'The astronomer announced that he had been mistaken. Earlier in the year he had predicted a meteorite shower.'

You not responded to this sentence as there was not a target phoneme.
Listen for the /k/ sound as in 'kite'. Ready?

> 'It appeared that a change of attitudes and biases was in the offing. The group of experts disagreed about desireable traits for a king in this day and age.'

Any questions? All right then, let's begin.

APPENDIX G

SOURCE TABLE FOR AMBIGUITY x CONTROL x GROUP

ANALYSIS OF VARIANCE

Source	Source of Squares	Df	Mean Square	Expected Mean Square	F
Mean	148.8384	1	148.8386	540.0	-
Groups	.3935	5	.0787	40.0	NS
Context	.0998	2	.0499	180.0	.01
Ambiguity	.0293	1	.0093	270.0	NS
Subjects (groups)	3.5676	84	.0425	6.0	-
G x C	.1481	10	.0148	30.0	NS
G x A	.5134	5	.1027	45.0	.01
C x A	.0686	2	.0343	60.0	.025
S x C(G)	1.5471	168	.0092	2.0	-
S x A(G)	.9459	84	.0113	3.0	-
G x C x A	.4229	10	.0423	5.0	.01
S x C x A(G)	1.4112	168	.0084	1.0	NS

APPENDIX H

SUMMARY OF RESULTS OF NEUMAN-KEULS TEST PERFORMED ON THE AMBIGUITY X CONTEXT INTERACTION

	A_2C_3	A_2C_1	A_1C_3	A_1C_2	A_2C_2	A_1C_1
A_2C_3					*	**
A_2C_1						*
A_1C_3						

KEY:

A_1 = Ambiguity

A_2 = Unambiguous Contol

C_1 = No Control

C_2 = Immediate Context

C_3 = Distant Context

* = $p < .05$

** = $p < .01$

APPENDIX I

DISTANCE (IN SYLLABLES) BETWEEN CONTEXT AND TEST ITEM

Sentence Number	Immediate Context	Distant Context	Sentence Number	Immediate Context	Distant Context
1	0	25	21	0	12
2	0	16	22	0	15
3	2	26	23	0	10
4	0	13	24	0	13
5	0	26	25	1	11
6	0	20	26	0	26
7	0	22	27	0	11
8	0	11	28	0	23
9	0	21	29	0	23
10	1	14	30	3	25
11	0	40			
12	0	14	Total =	7	552
13	0	20	mean =	0.23	18.4
14	0	23	range =	0-3	10-40
15	0	13	median =	0	16.5
16	0	11			
17	0	13			
18	0	23			
19	0	17			
20	0	15			

BIBLIOGRAPHY

Bever, T. G., Garrett, M. F. & Hurtig, R. The interaction of perceptual processes and ambiguous sentences. Memory & Cognition, 1973, 1, 277-286.

Bever, T. G. & Lackner, K. The underlying structures of sentences are the primary units of immediate speech processing. Perception & Psychophysics, 1969, 5, 225-234.

Cairns, H. Ambiguous sentence processing. Unpublished doctoral dissertation, University of Texas at Austin, 1970.

Cairns, H. The effects of bias on the processing and reprocessing of lexically ambiguous sentences. Journal of Experimental Psychology, 1973, 97, 337-343.

Cairns, H. Confirmation of the "Early decision hypothesis" and a comparison of psycholinguistic tasks. In preparation.

Caplan, D. Probe tests and sentence perception. Unpublished doctoral dissertation, MIT, 1971.

Conrad, C. Context effects in sentence comprehension: A study of the subjective lexicon. Memory & Cognition, 1974, 2, No. 1A, 130-138.

Cutler, A. & Foss, D. J. Comprehension of ambiguous sentences: The locus of context effects. Paper presented at the Annual Meeting of the Midwestern Psychological Association. Chicago, May, 1974.

Foss, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning & Verbal Behavior, 1969, 8, 457-462.

Foss, D. J. Some effects of ambiguity upon sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 699-706.

Foss, D. J., Bever, T. & Silver, M. The comprehension and verification of ambiguous sentences. *Perception & Psychophysics*, 1968, *4*, 304-306.

Foss, D. J., & Cairns, H. Some effects of memory limitation upon sentence comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, 1970, *9*, 541-547.

Foss, D. J. & Jenkins, C. Some effects of context on the comprehension of ambiguous sentences. *Journal of Verbal Learning & Verbal Behavior*, 1973, 12, 577-589.

Foss, D. J. & Lynch, R. Decision processes during sentence comprehension. Effects of surface structure on decision times. *Perception & Psychophysics*, 1967, *5*, 145-148.

Foss, D. J. & Swinney, D. On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning & Verbal Behavior*, 1973, *12*, 246-257.

Garrett, M. Does ambiguity complicate the perception of sentences? In *Advances in Psycholinguistics*, D'Arcais and Levelt (Eds.), North Holland Publishing Co., Amsterdam, 1970.

Hakes, D. Psychological Processes in Sentence Comprehension. Research Proposal submitted to National Science Foundation, June, 1971.

Hakes, D. On understanding sentences: In search of a theory of sentence comprehension. Final report submitted to U. S. Department of Health, Education, and Welfare, 1972.

Hakes, D. & Foss, D. Decision processes during sentence comprehension: Effects of surface structure reconsidered. *Perception & Psychophysics*. 1970, *8*, 413-416

Hebb, D. The organization of behavior. Wiley, N. Y., 1949.

Jarvella, R. & Herman, S. Clause structure of sentences and speech processing. Psychonomic Society, 1972, 11, 381-383.

Katz, J. T. & Postal, P. M. An integrated theory of linguistic description. MIT Press, Cambridge, 1964.

Kimball, J. Seven principles of surface structure parsing in natural language. Cognition, 1973, 2, 15-47.

Lackner, J. & Garrett, M. Resolving ambiguity: Effects of biasing context in the unattended ear. Cognition, 1972, 1, 359-372.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Maier, N. R. F. Reasoning in humans. Journal of Comparative Psychology, 1930, 10, 115-143.

MacKay, D. Aspects of the theory of comprehension, memory and attention. Quarterly Journal of Experimental Psychology, 1973, 25, 22-40.

Meyer, D. Dual memory-search of related and unrelated semantic categories. Paper presented at the annual meetings of the Eastern Psychological Association, New York, N. Y., April, 1971.

Meyer, D. & Schvaneveldt, R. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 1971, 90, 227-234.

Miller, G. A., Heise, G. A. and Lichten, W. The intelligibility of speech as a function of the context of the test material. Journal of Experimental Psychology, 1951, 41, 329-335.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Neisser, U. *Cognitive Psychology*, Appleton-Century-Crofts, N. Y., 1966.

Norman, D. Toward a theory of memory and attention. *Psychological Review*, 1968, *75*, 522-536.

Norman, D. & Rumelhart, D. A system for perception and memory. In *Models of Human Memory*, Norman (Ed.), Academic Press, New York, 1970.

Rubenstein, H., Garfield, L., & Millikan, J. A. Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, 1970, *9*, 487-494.

Rubenstein, H., Lewis, S., & Rubenstein, M. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning & Verbal Behavior*, 1971, *10*, 52-62.

Savin, H. B. & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, 1970, *9*, 295-302.

Thorndike, E., & Lorge, I. *The teacher's word book of 30,000 words*. Bureau of Publications, Teacher College. Columbia University, New York, 1963.

Treisman, A. M. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 1960, *12*, 242-248.

Weiner, B. J. *Statistical principles in experimental design*. N. Y., McGraw-Hill, 1962.

Woodworth, R. & Schlosberg, H. *Experimental Psychology*, Holt, Rinehart and Winston. N.Y., 1954.

VITA

David Allen Swinney was born in Indianapolis, Indiana on April 21, 1946, the son of Donald H. Swinney and Mary M. Swinney. He attended Uniondale High School in Uniondale, New York, graduating in 1964. Following this, he attended Indiana University, in Bloomington, Indiana, receiving the degree of Bachelor of Arts (with a major in Psychology) in June, 1968, and a Masters of Arts (with a major in Speech and Language Pathology) in August, 1969. From September, 1969, until August, 1970, he was employed as a research associate at the Center for Applied Linguistics, in Washington, D.C. In September, 1970, he entered the Graduate School at the University of Texas at Austin to work on a Doctor of Philosophy degree in Psychology.

Permanent address: 341 Lenox Avenue
Uniondale, New York 11553

This dissertation was typed by Edra Cassell and Arlene Walker.

APPENDIX IV

On the Understanding of Idiomatic Expressions

Linda L. Brannon

ON THE UNDERSTANDING OF
IDIOMATIC EXPRESSIONS

by

LINDA LORAINE BRANNON, A.B.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN
January, 1975

ACKNOWLEDGMENTS

Upon the successful completion of any dissertation, each writer heaves many sighs of relief, gives many shouts of joy and offers many thanks to those who helped. Of course, this writer is no exception. There have been many people who have offered suggestions, encouragement and sometimes just a pat on the head. All these were appreciated. But there are some people who deserve special thanks.

I would especially like to thank Dr. Michael L. Kamil who encouraged me to begin investigation into the problem of idioms. Without his interest and support I am sure I would not have attempted the investigation. His assistance with problems in Experiment I as well as his continued interest has been greatly appreciated.

I would also like to thank Dr. Donald J. Foss for his help with problems in design which arose in Experiment II.

Dr. Susan Schmerling has been most helpful in revealing the wisdoms of linguistics to me, and certainly deserves my thanks.

Another very special thanks goes to Dr. David T. Hakes who has helped with every aspect of this dissertation--from design problems to interpretation of the data. He has devoted much time and much energy to the problem of idioms, and I am most grateful for his concern and for his help.

This research has been supported largely by a grant (OEG-0-72-3941) from the National Institute of Education to Dr. David T. Hakes. I am most thankful to both the National Institute of Education and to Dr. Hakes for this support.

There are always many people other than teachers and advisors who deserve thanks, and rarely get it. Of my friends who have been interested in my research, my officemates--Judy Evans, David Swinney, Tom Bellhouse, Randy Bias and David Faye--deserve special thanks. They have, as the occasion demanded, diverted, comforted, encouraged, and (for the most part) shown extraordinary patience to me. I really do appreciate their continued kindness.

Barry Humphfus also deserves special thanks. Without his continued encouragement, faith and affection, I would have (to use phrases appropriate to the subject of this dissertation) climbed the walls, gone off the deep end and given up the ship long ago.

L.L.B. November, 1974

TABLE OF CONTENTS

Page

Acknowledgments iii

List of Tables and Figures vii

Chapter I -- INTRODUCTION 1

1.1. Psycholinguistics and ambiguity 2

1.1.1. Kinds of ambiguity 3

1.1.2. Experiments with ambiguity 4

1.2. Deviant sentences 6

1.2.2. Linguistics and ungrammatical sentences. . . 7

1.2.2. Psychology and deviant sentences 10

1.3. Idioms and linguistics 13

1.4. Psychology and idioms 20

Chapter II -- THE FIRST IDIOM EXPERIMENT 27

2.1. Pre-tests 30

2.1.1. Pre-test I 30

2.1.2. Pre-test II 31

2.2. Materials and design 32

2.3. Procedure 33

2.3.1. Subjects 35

2.4. Results and discussion 35

2.4.1. Analysis I 38

2.4.2. Analysis II 39

Chapter III -- THE SECOND IDIOM EXPERIMENT--THE AMBIGUITY VARIABLE 41

3.1. Design and materials 43

3.1.1. Pre-test 48

3.2. Procedure 48

3.2.1. Subjects 49

3.3. Results and discussion 51

3.4. Bias 57

3.4.1. Materials and design 58

3.4.2. Procedure 59

3.4.2.1. Subjects 59

3.4.3. Results and discussion 60

Chapter IV -- CONSIDERING THE VARIABLE OF GRAMMATICALITY 65

4.1. Design and materials 67

4.2. Pre-test 69

4.3. Procedure 71

4.3.1. Subjects 72

4.4. Results and discussion 73

Chapter V -- A MODEL TO EXPLAIN THE DATA 78

APPENDICES 96

BIBLIOGRAPHY 119

LIST OF TABLES AND FIGURES

Page

TABLES

1. EXAMPLE MATERIALS AND DESIGN FROM EXPERIMENT I 34
2. ERROR RATES FOR EXPERIMENT I 37
3. RESULTS FROM EXPERIMENT I 37
4. AN EXMAPLE OF MATERIALS USED IN EXPERIMENT II 47
5. RATINGS ON PRE-TEST FOR EXPERIMENT II 50
6. ERROR RATES FROM EXPERIMENT II 52
7. RESULTS FROM EXPERIMENT II 53
8. CORRELATIONS WITH BIAS 62
9. EXAMPLES OF MATERIALS USED IN EXPERIMENT III 68
10. PRE-TEST RATINGS FROM EXPERIMENT III 70
11. RESULTS FROM EXPERIMENT III 74
12. ERROR RATES FROM EXPERIMENT III 74
13. NUMBER OF POSSIBLE READINGS FOR MATERIALS FROM EXPERIMENT III 83
14. NUMBER OF POSSIBLE READINGS FOR MATERIALS FROM EXPERIMENT II 83
15. AVERAGE NUMBER OF COMPARISONS NECESSARY FOR A MATCH 84
16. AVERAGE NUMBER OF COMPARISONS NECESSARY FOR A MATCH MATERIALS FROM EXPERIMENT I 84

FIGURE

1. A SKETCH OF THE DESIGN USED IN EXPERIMENT II 45

CHAPTER I

INTRODUCTION

The field of psycholinguistics, that area of psychology which is concerned with the cognitive processes accompanying language, is a new area of concern relative to the field of experimental psychology. One of the problems of interest to psycholinguists is the comprehension of language. This was one of the first facets of language to be experimentally investigated, and is the area in which the most research has been done.

Comprehension researchers have investigated many factors which affect language comprehension, including ambiguity, complex verb structures, relative clause reduction, self-embedding, grammatical deviance and other factors which could affect the comprehension process. Each of these variables has been studied separately, but there is a phenomenon in which two variables, ambiguity and grammatical deviance, converge: the phenomenon of idiomaticity, which has been largely neglected by researchers in the field of psycholinguistics.

Idioms are extremely common in everyday language use. An idiom can be loosely defined as a string of words whose meaning is not a compositional function of the meanings of the words in the string. Idiomatic expeessions are of two types: those with and those without alternative literal interpretations. Those with an alternative literal interpretation are ambiguous, because they can be interpreted either as an idiom or as a normal constituent. For example, KICK THE BUCKET is ambiguous; it has, for most people, two completely different and unrelated meanings. Idioms of the other type, those without alternative literal interpretations, are not ambiguous, but are, in the strict, formal sense of grammar, deviant. For example, HE DIDN'T STAND A CHANCE can be an idiom or it can be gibberish since the verb STAND is either intransitive, having no object, or requires an adverbial.

Both idioms with and those without alternative literal interpretations are phrasal idioms which consist of more than two words and seem to have the grammatical structure of a phrase. There are also lexical idioms which consist of two words only and seem to be lexical items, for example, RED HERRING, whose meaning has nothing to do with either a color of the spectrum or with fish.

Linguists have studied the peculiarities of idioms to some extent. Idioms are of interest to linguists not only because of their frequency in language use, but also because they are a puzzling linguistic phenomenon: they either are ambiguous or they are ungrammatical. Both types of idioms exhibit transformational peculiarities, another peculiarity of idioms which has been of interest to linguists. Psycholinguists have studied both the comprehension of grammatically deviant sentences and the comprehension of ambiguous sentences, but have largely failed to see idioms as an area of possible convergence for these two variables. Idioms could provide research possibilities for both those variables as well as being an interesting phycholinguistic phenomenon in itself.

1.1. Psycholinguistics and ambiguity

One prominent approach to the problem of comprehension has been through the study of ambiguity and how it affects the process of comprehension. Ambiguity is a common occurrence in everyday experience for any listener, and even those sentences which are not ambiguous in their entirety are often, at some point, ambiguous (see, e.g., Cairns, 1970). For this reason, an understanding of the ways in which ambiguity affects the processes of comprehending sentences is generally regarded as revealing and important to the understanding of language comprehension in general.

1.1.1. Kinds of ambiguity

Linguists regard ambiguity as being of several types. A sentence can be ambiguous because of grammatical structure, because a single lexical item has two possible meanings in the sentence, or because a word has two possible pronunciations. Sentence (1) is ambiguous if it is spoken because the word sale sounds like the word sail. Sentence (2) is ambiguous because the word bat has two possible meanings in the sentence, either a) a piece of wood with which one can hit a ball or b) a small flying nocturnal mammal.

(1). The $\left\{\begin{matrix}\text{sale}\\\text{sail}\end{matrix}\right\}$ was very large.

(2). My little brother found a bat in the attic.

The grammatical structure of a sentence can create ambiguity by way of its constituent structure or by the underlying relationships of the lexical items. Both sentence (3) and sentence (4) are ambiguous because of their grammatical structure. Sentence (3) is ambiguous because of the two different possible groupings of the words old men and women. This sentence could mean a) old men and old women or b) old men and also women (who are not necessarily old). There are no superficial groupings which will resolve the ambiguity in sentence (4). Sentence (4) is ambiguous because of the underlying grammatical relationships between the words. The

(3). There were old men and women at the store.

(4). Visiting relatives can be boring.

sentence can mean either a) for one to visit one's relatives can be boring, or b) for one's relatives to visit can be a boring experience. The structure of the sentence leaves unclear who is doing the visiting.

1.1.2. Experiments with ambiguity

The problem of the comprehension of ambiguous expressions has been given quite extensive experimental investigation. The methods of investigation have varied, and so have the results. And yet, it does seem possible to find some coherent, consistent explanation of the apparently conflicting data.

A number of investigators (MacKay, 1966; MacKay & Bever, 1967; Foss, Bever & Silver, 1968; Cairns, 1973) have found differences in the comprehension of lexical ambiguities and underlying structure ambiguities. In some cases (MacKay, 1966; MacKay & Bever, 1967; Foss, 1970), the lexical ambiguities seem to be easier to deal with in comprehension tasks than underlying structure ambiguities. Yet in other tasks (Foss, Bever & Silver, 1968; Cairns, 1973) lexical ambiguities seem to be the only kind that affect comprehension. Several attempts have been made to tell a coherent story about the comprehension of ambiguous sentences (Cairns, 1970; Garrett, 1970; Bever, Garrett & Hurtig, 1973). These explanations point out that the differing data were obtained using different tasks. MacKay (1966) used 'time to discover ambiguity' as a measure of processing time; MacKay and Bever (1967) used a sentence completion task; Foss, Bever and Silver (1968) used a picture verification task; Foss (1970) used a phoneme monitor task; Garrett (1970) used a 'click' location task; and Cairns (1970, 1973) used a 'decision about compatibility of sentences' measure. The experimenters who used a task which measured processing complexity during the comprehension process (Foss, Garrett) have found that lexical ambiguities cause an increase in processing immediately following the presentation of the ambiguity. The lack of an ambiguity effect later in the sentence or after the completion of the sentence (Bever, Garrett & Hurtig, MacKay & Bever, Cairns, in preparation) indicate that this effect of processing increase is rapidly handled by the comprehension mechanism in such a way that the difficulty is resolved very quickly. These same experiments suggest that the underlying structural ambiguities are handled more slowly, and that structural ambiguities are not resolved as early, perhaps not until a clause boundary has been crossed (Cairns, 1970). These results do indicate that during the comprehension process both readings of an ambiguity are available and that some decision between readings is made with varying degrees of ease. It should be noted that disambiguiting contexts can have some effect on this decision process (as shown by Foss, Bever & Silver; Cairns; Carey, Mehler & Bever), but a decision for one or the other reading is made even in the case of no disambiguiting context.

Bever, Garrett and Hurtig (1973) have created a model for sentence processing in which the clause unit plays a very prominent role. This model suggests that lexical items and structural relations are computed within each clause. This explains the temporary and immediate processing load increase caused by ambiguous lexical items. After each clause, some kind of processing occurs which recodes the clause from grammatical relations plus lexical readings into a more abstract form in which these grammatical relationships are not as readily available as during the initial processing of the clause. Since a recoding of the grammatical relations takes place at this point, the time immediately after the clause has been completed is the time during which underlying structure ambiguity would become apparent and be resolved. This model does seem to be a more coherent explanation of the data than has previously been proposed.

But where do idioms fit into this picture? Do ambiguous idioms correspond more closely to lexical or structural ambiguities? A case can be made for either. If idioms are, in our 'mental lexicons' amalgamated into long 'words' as suggested by Chomsky (1965), then an idiom could be processed the same as an ambiguous lexical item. If the structure is retained in idioms, then it is possible that idiom processing proceeds more like that of structural ambiguities. Or perhaps it is not accurate to establish a paralled with either type of ambiguity. Perhaps idioms are accompanied by processing of a special type which is not closely analogous to either lexical or structural ambiguities.

1.2. Deviant sentences

Sentences which are not completely correct from the standpoint of formal grammar (à la Chomsky) have also been of interest to those investigating comprehension. Such sentences can either be complete gibberish, or simply lacking the strict well-formedness which would make them acceptable to a junior high school grammar teacher. Sentence (5) is of the former type and (6) is of the latter. One would be hard-pressed

(5). The go of for box me.

(6). To the movies.

to invent a situation where (5) would be acceptable and sensible (though such context probably could be constructed with a little creativity), but (6) is a perfectly acceptable elliptical response to a question such as "Where are you going this evening?" This last example points out the importance of these types of constructions: they are extremely common to everyone in ordinary language use. Indeed, one might be more surprised to hear the full sentence reply, "This evening I am going to the movies," than the partial sentence.

As has been observed, in ordinary usage, ungrammatical sentences are far more common than strictly grammatical ones (Chomsky, 1957). Sentences which are ungrammatical can be like Sentence (6), incomplete fragments of sentences, or they can be deviant for many other reasons; for example, the lack of number agreement between subject and verb, an improperly conjugated verb, or an improperly appended clause. One need only listen to oneself and to the language of others to substantiate this observation.

1.2.1. Linguistics and ungrammatical sentences

Much of the work done in linguistics in the past decade has been an attempt to characterize speakers' knowledge of their language. Many of the attempted descriptions are in the transformational framework. In a transformational grammar as proposed by Chomsky (1965), an adequate model of language would characterize all and only the sentences of a language which the speakers of that language consider well-formed. That is, if the model for a language is adequate, it will characterize no ungrammatical sentences.

Even if such a model for grammar were complete (and to date, none is), there would be some problems with such a description: that model would not be complete with respect to any speaker's knowledge of his/her language if it contained only knowledge of the grammatical sentences, since much that is known and commonly used is not in terms of strictly well-formed sentences. This problem has been recognized by several linguists, and Katz (1964) states the problem by saying

> As everyone knows, speakers can understand not only well-formed utterances of their language (sentences), but they can also understand utterances that are not well-formed, such ungrammatical strings, semi-sentences as we shall call them, are frequently heard in everyday conversation and are, in certain circumstances, the best way of communicating exactly what one wishes. . . . (p. 400)

Several attempts have been made to explain this capacity within the framework of a transformational grammar. Katz (1964) formulates an explanation which extends the capacity to deal with the well-formed sentences to the ones which are not well-formed, allowing comprehension to proceed by rules which establish a connection to well-formed sentences. The ill-formed sentences are understood because they have well-formed sequences and the grammar contains rules which connect the ill-formed sentences to the well-formed counterpart using the well-formed sequences as the device to accomplish this link.

Chomsky (1964) also has a proposal for dealing with any ungrammatical sentences:

> . . . we attempt to impose an interpretation on it, exploring whatever features of grammatical structure it preserves and whatever analogies we can construct with perfectly well-formed utterances. (p. 389)

Chomsky establishes a system which assigns degrees of grammaticalness which measures the remoteness of utterances from the set of well-formed sentences generated by the transformational grammar. This measure of remoteness is a way to set up a hierarchy of acceptability for deviant utterances. Katz criticizes this description because he says it fails to set up any division between those deviant utterances which are not intelligible and those which are, something which his system does and which he finds necessary to explain the intelligibility of some deviant utterances.

Generative semanticists have a different view of grammaticality. Most generative semanticists view well-formedness of a string as relative to the context in which it appears, and also to the knowledge and assumptions the speaker has about the world. Lakoff (1971) argues that one cannot speak of the well-formedness of lack of well-formedness of any sentence in isolation. For him, presupposistions that the speaker has about the nature of the world determines the well-formedness of any sentence for any speaker. He does, however, admit that there are certain constraints on sentences which, when violated, can cause absolute ungrammaticality. For example, sentence (7) is, even for Lakoff, absolutely ungrammatical.

(7). * I went Boston to.

His concern is not with this kind of absolute ungrammaticality, but rather with the kind of ungrammaticality which results when a speaker uses language which violates his presuppositions about the nature of the world. For example, Lakoff would surely consider sentence (8) (in the literal sense) ungrammatical, assuming that Bill is human and that humans cannot fly without assistance.

(8). * Bill flew off the handle.

Looking to linguistics, there is no united opinion of ungrammaticality and the lack thereof. What constitutes well-and ill-formedness varies with linguistic theories. The generative semanticists view of grammaticality would seem to be the more promising one with respect to psychology. Since this view incorporates information about the world into its view of grammaticality, it would seem to be the kind of theory that could be expanded into some kind of psycholinguistic theory of how people handle deviant sentences. But as yet Lakoff's approach remains little more than a suggestion, and it remains unclear whether this promise will be realizable.

1.2.2. Psychology and deviant sentences

Even though the occurrence of deviant sentences is commonplace, investigations into the comprehension of sentences which violate the rules of formal grammar have been quite sparse. Downey and Hakes (1968) investigated this problem by using three tasks: an acceptability rating task, a recall learning task and a paraphrase task. As materials they used sentences which, in a Chomskian generative transformational-type grammar, contained either no rule violation, a violation of a phrase structure rule (such as using a noun in place of a verb), a violation of a subcategorization rule (such as using an intransitive verb where a transitive one should be used), or a violation of selectional restrictions (such as using a verb which requires a human subject with a subject noun which is not human). In a transformational framework, these kinds of rules are hierarchical; that is, if a phrase-structure rule is violated, then there is also a violation of subcategorization and selectional restrictions; if there is a violation of subcategorizational rules, then there is also a violation of selectional restrictions but not necessarily of any phrase-structure rules. One could, therefore, predict from this system that a violation of phrase-structure rules would be a more serious violation of the comprehension ease than a violation of either subcategorization or selectional restrictions, and that subcategorization violations would be more serious than selectional restriction violations. These were the predictions made by Downey and Hakes. The results of the acceptability rating task were exactly as predicted, but the results of the other two tasks varied from these predictions. In both the learning and paraphrase tasks, the sentences with subcategorization violations appeared to be easier to deal with than those with selectional restriction violations. This result is contrary to the predictions made according to transformational model, and can possibly be explained by examining the strategies used by the subjects to do the tasks. It is at least clear that deviant sentences seem to be harder to deal with for the language user than non-deviant ones.

Danks (1968) has also investigated the comprehension of deviant sentences using tasks that may be closer to the normal comprehension process than those used by Downey and Hakes. Danks performed two experiments. In one experiment he collected ratings on comprehensibility of a group of sentences which varied in several respects. He constructed sentences which were grammatical and meaningful, ones which were ungrammatical but meaningful, ones which were non-meaningful but grammatical, and a group of sentences which were both ungrammatical and non-meaningful. His sentences which were constructed to be non-meaningful but still grammatical, actually contained many violations of selectional restrictions. The sentences he considered ungrammatical were "word salad", the words of a grammatical sentence thrown in pseudo-random order with none of the original grammatical relations preserved. His variables of grammaticality, meaningfulness, and the interaction of these accounted for the vast majority of variability for this task. He used this set of materials for his second experiment in which he had subjects perform one of three tasks. Subjects either had to "understand" the sentence, "correct the grammar" or "correct the meaning" of the sentence. The time required to perform the task was the dependent variable. In all the tasks, meaning was the most important factor affecting reaction time, with grammar also having some effect. Keeping in mind that grammaticality and meaningfulness are not really separable (Downey & Hakes, 1968) for the materials used in Danks' experiments, it is not possible to assign grammaticality the lesser role. It is possible to conclude from Danks' experiments that deviant sentences do cause difficulties in comprehension, and that selectional restriction violations which cause confusion as to the meaning of the sentences can be a significant factor in comprehension.

Here, one point about the processing of ungrammatical sentences should be made, and the implications of that point discussed. It must be assumed on grounds of logic that ungrammatical sentences are identified as ungrammarical by the attempt of the comprehension mechanism to comprehend these as grammatical sentences and failing. This must be assumed because it is not possible to identify the lack of grammaticality until some kind of processing has been done, and that processing can only be initially that of normal comprehension. This implies something very important for the understanding of the mechanism involved. Since the attempt has been made to process the ungrammatical sentence in a normal way, the process of comprehending an ungrammatical sentence must involve some process in addition to the normal, and not just a different process from the normal. It should also be stated that no one knows what this process is at this point. The Katz model (discussed in Section 1.2.1.) is inadequate because for most ungrammatical sentences, there is no one P-marker which represents the 'closest underlying P-marker' that the Katz model postulates as the mechanism for obtaining a meaning for ungrammatical sentences. For any ungrammatical sentence, there are several possible P-markers which could represent the grammatical structure from which such a string could come. However, unless the unique P-marker is known, the path to any such P-marker is not made clear by such a process. The Chomsky description of the problems involved is not sufficiently detailed to be considered a model for comprehension. Including the kind of information that Lakoff argues as being important to grammaticality seems promising for development into a model, but the promise is all that yet exists.

1.3. Idioms and linguistics

Idioms, representing the possible convergence of the variables of grammatical deviance and ambiguity, have been provocative to linguists because of the problems that the description of idioms brings for almost any proposal for language description. It is immediately obvious that idioms are syntactically and/or semantically peculiar. First, the meaning of the string of words in an idiom is not a combination of the meaning of its compositional parts as is generally thought to be the case for normal sentences. For example, KICK THE BUCKET does not have the meaning of KICK + DEFINITE ARTICLE + BUCKET, but rather this string of words has a meaning very similar to die.

Chomsky (1965) has proposed treating idioms, essentially, as long words with no internal structure. This 'unit' treatment of idioms has several problems. If an idiom were one long 'word' (all part of one complex symbol in deep structure, in the Chomsky terminology), then it would be impossible to account for the fact that the morphological units in many idioms (roughly, words) act much the same as those morphological units when they are not in idioms. For example, the verb KICK in KICK THE BUCKET becomes KICKED in the past tense, KICKED THE BUCKET just as it does in KICKED THE BALL, and not KICK THE BUCKETED as would be expected if the idiom were a unit. The unit treatment does not allow any connections (semantic or syntactic) to be made between the two verbs SHOOT in SHOOT THE BREEZE and SHOOT THE GUN, nor does it allow for the fact that these two both have the same irregular past tense form, SHOT. Under the unit treatment these similarities would be chance occurrences. The unit treatment would allow a lexical item (the idiom) to replace lexical items which do not form a constituent, a thing which, Newmeyer (1973) states, does not happen otherwise.

Idioms themselves need not be constituents (a peculiar thing if they are considered to be lexical items), nor do they necessarily replace constituents. The internal structure of idioms becomes apparent when some of the parts (words) in the idiom are moved to other parts of the sentence, breaking up the idiom, yet the idiomatic meaning is retained. For example, both sentences in (9) retain the idiomatic meaning, even though b) is passive and the idiom is broken into two segments.

(9). a) John broke the ice with his comedy routine.

b) The ice was broken when John did his comedy routine.

All of these facts about idioms indicate that the Chomsky-type unit formulation is not an adequate way to characterize idioms in a linguistic description.

There have been two other attempts to characterize idioms within a transformational grammar framework. Katz and Postal (1963) agreed that a unit treatment for idioms would cause problems for a grammar. They proposed a system which contains a special phrase idiom dictionary in which the syntactic and semantic information about idioms is stored. When an idiom is retrieved from this storage, the idiomatic reading is associated with a constituent which dominates the idiom and not the items in the idiom. The literal counterpart of an idiom is different in that the meanings for the items of such a phrase are associated with the individual items, and not the entire 'phrase idiom'. Therefore, the derivation of the meaning of a literal phrase is different from that of an idiom. This system cannot handle idioms which have no literal counterparts, those which are, strictly, ungrammatical. Because they have no literal counterparts, they would be ungrammatical in this sort of transformational framework, and therefore, unallowable structures. This poses a great problem for this type of model of grammar: ill-formed structures are not within the realm of language since they are not produced by the grammar, yet these kinds of structures do seem necessary because many idioms are of this type. Only with an idiomatic reading can these structures have any interpretation. Any proposed model of language would have to generate structures which are ungrammatical for these idioms with no literal interpretation and also be able to assign either an idiomatic reading or a non-idiomatic reading to those which are syntactically well-formed and therefore, ambiguous. Katz and Postal's system could do the latter, but not the former. They recognize that their description is inadequate in this way, but believe that such grammatically deviant idioms are handled by the same processes that enable people to interpret other ungrammatical but intelligible phrases (See section 1.2.). Since the meaning of an idiom is not a compositional function of its parts, it is difficult to see how an idiomatic meaning could be obtained by this method of analogy to the closest grammatical construction; they do not explain in any further detail how this could be possible.

In Chomsky's system, a complex symbol is roughly equivalent to a word, or, more accurately, equivalent to the abstraction that corresponds to a word in the syntactic description of language. These abstractions have no linear internal structure in the syntactic part of the grammar. In Chomsky's unit treatment, where an idiom is equivalent to a single complex symbol, the idiom has no internal structure. In Fraser's description of idioms (1970), it is possible to have a single branch of a tree structure end with a string of complex symbols rather than just one All other types of words are represented by only one complex symbol in normal constructions, just as in Chomsky's system.

Fraser (as well as other linguists) claims that idioms all exhibit some transformational peculiarities. That is, phrases that are idioms cannot always have the same transformations applied to them that non-idiom phrases can. In the case of many ambiguous idioms, when a transformation is applied, the idiomatic meaning is no longer present. For Example, (10a) is ambiguous but (10b) is not (for many speakers).

(10). a) They were shooting the bull.

b) It was the bull that they were shooting.

Fraser believes that the behavior of idioms with respect to transformations is largely unpredictable and appears to be idiosyncratic to particular idioms and even to particular speakers.

Fraser tried to explain the transformational peculiarities of idioms by setting up a hierarchy of frozenness for idioms which ranks idioms according to the transformations which can be performed on them with a retention of idiomaticity. According to this ranking, any idiom which can undergo certain kinds of transformations is placed at a particular level and this guarantees that it can undergo the transformations at any lower level in the hierarchy. The individual idioms within the hierarchy can vary from person to person, but a hierarchy does exist for each person and is internally consistent with respect to the placement of idioms.

The hierarchy was intended to be an explanation of idioms within the Chomskian framework, but that framework is stretched considerably by this explanation. The hierarchy never reaches the stage of actual explanation but seems only a description of the characteristics and perhaps merely the structural complexity of idioms. For Fraser an idiom must at least be analyzable as verb and direct object before it can fit into the most flexible level of the hierarchy. This restriction would exclude many idioms from the most flexible level of the hierarchy simply because they do not conform to a certain structural description. If an idiom must conform to these structural descriptions, then it cannot be said that the hierarchy is actually descriptive of anything more than the characteristics of some individual idioms. This may be a very interesting list to have, but it does not explain why the transformational peculiarities of idioms exist, nor does it establish any relationship between individual idioms and their transformational peculiarities, and so fails in its stated purpose since idiom behavior is still idiosyncratical with respect to transformations.

There have also been attempts to fit idioms into a semantic model of language description. Weinreich (1966) proposed a method to handle lexical idioms, such as HOT POTATO, RED HERRING, etc., by breaking such idioms down into component parts which would be listed in some mental dictionary along with a description of the context into which each part must fit. For phrasal idioms he proposed an idiom list, each entry of which contains the string of words in the idiom together with semantic and syntactic information necessary to their proper use. He sets up the definitions in his system such that those strings of words which do not have literal counterparts are not idioms, and he says that his system does not apply to such strings. Although he did acknowledge that such strings should be included in an adequate description of language, his explanation was not so ambitious as to include these.

Chafe (1968) criticized both the Katz-Postal and the Weinreich method of handling the description of idioms. He asserted that while the Katz-Postal system would allow for the idiomatic interpretation of a phrase, it would not account for all of the transformational peculiarities that idioms have. Chafe also criticized Weinreich's explanation of idioms because Weinreich had to include the transformational peculiarities of each idiom as part of each entry in the idiom list. This is not adequate, Chafe argued, because it relegates the information to the status of random information, which he believes that it is not. He also criticized Weinreich for not including idioms which are not syntactically well-formed in his description, saying that putting them into a completely different part of the language system characterized them as totally different phenomena, which they are not.

In Chafe's description of idioms (and language in general), semantic information is of primary importance. He views language as a conversion of semantic units to phonetic units, and views idioms as products of a historical process by which certain strings of words undergo a semantic split such that a string retains its syntactic arrangement but is given a different semantic interpretation. After a while, the semantic interpretation associated with the original syntactic arrangement (literal interpretation) becomes less prominent in the language, but the new semantic interpretation remains in the form of the original syntactic arrangement. He hypothesizes literalization rules which recapitulate this historical process in a reverse direction; a semantic interpretation becomes a semantic unit which requires conversion into an arrangement of other semantic units before it can be encoded into sound. This is the way that idioms come to exist in the language and also the way they function in the language after the historical process of their creation has been achieved. This explanation would, of course, account for the fact that idioms have a meaning other than the meaning of their compositional parts, and would also account for their transformational defectiveness in an interesting way. Since idioms are, in Chafe's formulation, single semantic units, they would have the grammatical properties of those semantic units and not those of their surface structures. Such an explanation can also account for the syntactic ill-formedness of idioms by saying that they were not always so, but came to be ill-formed when the syntactic rules of the language changed. This is a very interesting explanation, and may prove to be an adequate account of the evolution of the idioms themselves, but is not illuminating as to how the idioms are understood by speakers of the language.

Newmeyer's (1973) explanation of idioms is similar to several others. He hypothesized an idiom list of semantic forms and rules by which the list would connect to the surface structures. This formulation would insure that all idioms have two meanings, one of which would be realized as an idiom, the other as a non-idiom, that is, a literal counterpart of the idiom. This leaves his formulation lacking an explanation of ill-formed idioms, and he acknowledged that such are costly for his theory.

Most of the explanations that linguists have put forth contain some kind of special listing for idiomatic expressions which serves to connect the idiomatic non-compositional meaning to the idiom structure, and which allows for both an idiomatic and a literal interpretation of the phrase. As previously mentioned, this causes a problem with the idiom structures which have no literal counterpart, and the linguists handle this problem poorly, either with additional complication to their respective theories or by denial and dismissal of the relevence of the problem. The transformational type theories seem to do more poorly than the semantic type theories in dealing with the transformational peculiarities of idioms. Since the semanticists consider meaning the determining factor, their explanations of idiom properties being determined by meaning seem more explanatory than the descriptions of the transformational idiosyncracies put forth by transformationalists, which have no coherent motivation. Newmeyer's description, which incorporates both syntactic and semantic information, is perhaps the best of both worlds, although it also suffers from extra complication in describing the ill-formed idioms. The linguists have thought about the problems that idioms make for language description at some length. It is obviously a difficult and frustrating problem. However, psycholinguists interested in the comprehension of language are faced with different problems when studying idioms.

1.4. Psychology and idioms

How are idioms processed psychologically? Are there special comprehension procedures that accompany the decoding of an idiom? The linguistic descriptions could suggest either greater or lesser complexity associated with an idiomatic structure. If an idiom were a long 'word', the implication would seem to be that it has less than normal structure, and a less complex structure could mean that 1) idioms are easier to understand because it is easier to assign a meaning to a less complex structure, or 2) idioms are harder to understand because the lack of structure is a complication for obtaining a meaning. Either way, this might be reflected in the comprehension process. Other linguistic descriptions suggest a special list of idiom entries and a special process in the language for idioms. This would suggest a special 'idiom mode' of comprehension which could be reflected in psychological processing.

If such an 'idiom mode' does exist, then the question of how this mode is triggered arises. What is necessary for the comprehension device to switch from 'normal processing mode' to 'idiom processing mode'? For an ungrammatical idiom, one immediately evident possibility is that when the comprehension device obtains the meanings of the words plus the grammatical relations and finds that the assembled meaning is anomalous, some special 'idiom mode' of comprehension is triggered. The question more difficult for the ambiguous idioms. When an underlying meaning is obtained for such a string, a non-anomalous meaning does exist. Perhaps it is only through context that the anomalous nature of this meaning is determined and a re-computation is required. Or perhaps the 'idiom mode' is triggered by the appearance of a key word in the idiom (of ungrammatical or ambiguous type), and the ambiguity is resolved much the same as any other ambiguity if the triggering causes ambiguity. The different possibilities suggest predictions which could be resolved by research. But is there any evidence for an 'idiom mode'?

Bobrow and Bell (1973) hypothesized that there does exist a special idiom mode of processing distinct from the literal mode used for other sentences. In order to substantiate this, they used a 'set' procedure

(Carey, Mehler & Bever, 1970) in which they had two groups of subjects read five sentences. In the literal set group, the first four of the sentences were non-idiomatic surface structure ambiguities, and in the idiom set group the first four sentences were ambiguous idioms. In both groups the last (test) sentence was an ambiguous idiom. Each sentence was printed on the page of a booklet, and its two possible meanings were printed on the reverse side of the page. The subjects were to read each sentence, discover the two meanings, and then turn the page and mark which of the two meanings they obtained first. The first sentence in the idiom set was used as a baseline response to idioms, from which the proportion of subjects detecting the idiomatic reading first was calculated. This proportion was compared to the proportion from the last (test) sentence from each group. The baseline response was .82, that is 82% of the subjects originally obtained the idiom reading first when they had no previous context, that is, when they had no 'set'. The proportion of subjects who obtained the idiomatic reading first after the literal set was .57, and after the idiom set the proportion was .84. These two are significantly different. However, the proportion for the test sentence in the idiom set is not different from the baseline for that group. Bobrow and Bell suggested that the failure to find an inflation of the 'obtained idiom first' response was due to a ceiling effect--the .82 is a high response rate initially. They unfortunately did not pre-test their materials, so it is not known whether this is the general explanation or is the case just for the limited set of idioms that they used. This failure to pre-test was a serious oversight considering the amount of individual variation that occurs with the familiarity of idioms. Because of their failure to pre-test they did not know that the idioms they used would be initially so highly biased toward the idiom reading as to prevent (so they say) a further inflation of the idiom response by their experimental procedure. Because they did not pre-test, they do not know whether an increase would have even been possible since it is possible that some of the idioms were not familiar to some subjects. This would mean that some subjects would never give an idiom reading because they do not know an idiom reading for that sentence. If this were the case, not only would it weaken the possible inflation of 'obtain idiom reading first' but it would also weaken their 'set' procedure: seeing sentences which were both idiomatic and non-idiomatic as a set would not be nearly as strong as seeing all idioms.

Bobrow and Bell conducted a second experiment to clarify the results of their first; to eliminate the possible ceiling effect from their materials, and also to learn whether the processing of idioms corresponds more to that of lexical ambiguities or more to that of structural ambiguities. Their second experiment was very similar to their first, except that there were two groups of literal set conditions, one with lexical ambiguities, and one with surface structure ambiguities created to test the similarity of processing between idioms and lexical ambiguities. If the idiom received a higher number of idiom readings after either the set of lexical ambiguities or the set of structural ambiguities, then it would be possible to say that the facilitation occurred because of the similarity in processing. The idiom set groups received initially a low-bias idiom which served as a baseline, four high-bias idioms which served as the set and one low-bias idiom which was the test. The lexical and surface structure ambiguity groups received five of the appropriate sort of ambiguities and then a low-bias idiom test sentence. In this experiment, the idiom set did increase the proportion of 'obtained idiom reading first' responses over the baseline, and the baseline was more in line with the literal set groups than in their previous experiment. The experiment provided no support for the hypothesis that idiom processing is more closely related to lexical ambiguity since both the lexical and structural ambiguity groups were equal. The experiment did provide support for the hypothesis that there is different processing with idiomatic material.

Bobrow and Bell's experiments can be criticized on several counts. Their failure to pre-test their materials was unfortunate; this failure could have caused the failure of the first experiment and did cause the loss of one of the test sentences for the second. Using low-bias idioms as they did in their second experiment, it is even more likely that some subjects were not familiar with all of the idioms. As previously mentioned, this considerably weakens their set procedure and decreases the probability of obtaining the predicted effect. Since the effect was obtained in their second experiment despite these oversights in procedure, the effect must be a substantial one.

It is questionable whether the dependent variable, self-reports of which meaning was seen first, is a very accurate and revealing one; language processing happens so quickly that introspection is often an inaccurate measure of processing procedures.

Bobrow and Bell discuss the comprehension of idiomatic expressions in terms of

> Discovery of the idiomatic meaning of an idiom seems to result from processing the idiom as a word. Such a processing strategy differs from that which has been suggested for literal expressions, wherein each word is perceived, its meaning discovered, and then mapped into a semantic whole such that the meanings of the individual words relate to produce a meaning for the sentence. (p. 344)

They believe that this implies a dichotomy of processing modes for which they found support. Their support for a dichotomy of processing modes can be questioned. They undoubtedly found some difference in the processing of idioms and normal sentences, but this difference does not necessitate a dichotomy. Mehler and Carey (1968) using the same kind of set procedure found a difference in the processing of progressive and adjectival structures of the type THEY ARE VERBING NOUNS, but they did not claim that there is a dichotomy of processing between these two kinds of structures. The claim of a dichotomy of processing is stronger than can be claimed from this type of experimental procedure. And even if a dichotomy of processing was found, it would not imply that idioms are processed as words, but merely that something about the processing of idioms is different from that of literal expressions. Their suggestion that

> . . . combining several words into a complex 'idiom word' and finding the meaning of the phrase by a search through a mental 'idiom word' dictionary (p. 343)

has the connotations of the 'unit treatment' which has been often disputed, is not supported by the notion of two modes of processing, and would be indistinguishable, on the evidence of their results, from a model which combines words into an idiom phrase which is decoded by searching through a mental 'idiom list'. The complexity of idiom comprehension is a very interesting one, the knowledge of which is vital to any statement of idiom processing. Neither of the Bobrow and Bell experiments has any evidence on the complexity of idiom comprehension process.

Before attacking the problem of the complexity of idiom processing the factors possibly affecting idiom comprehension should be reviewed. First, are idiom sentences processed differently than other kinds of sentences? There is some indication from Bobrow and Bell (1973) that they are, but there is no indication of their comprehension complexity. This, then, should be investigated first; if there is no indication that sentences containing idioms are different from other kinds of sentences, then the other questions are not relevant. If some difference in complexity of comprehension is found for idioms, this leads to other questions. For example, what effects do the different types of idioms have on the comprehension process? Are ambiguous idioms processed in the same way as other ambiguities? Are they processed differently from unambiguous idioms? Are the unambiguous idioms processed like other ungrammatical strings? These questions suggest a series of experiments on the nature of idiom processing, but the first question is: Are idioms different?

CHAPTER II

THE FIRST IDIOM EXPERIMENT

Are idioms more complex for the comprehension device than normal sentences? If idioms are more complex psychologically, then they should be harder to deal with in comprehension tasks than normal sentences. If some sort of amalgamation of meaning has taken place, then idioms could be easier to deal with than normal sentences: and if the comprehension of idiomatic expressions proceeds the same as that of other sentences, there should be no difference in performance on comprehension tasks. As an initial attack on the problem of the complexity of idiom comprehension, an experiment was designed to examine the various possibilities of comprehension complexity.

The experiment measured processing complexity by using a reaction time measure. The amount of time which was taken to decide on the similarity of meaning of two sentences both presented visually was indexed by a button press response. The reaction measure would include the time necessary to comprehend both sentences as well as the time necessary to make the decision about the similarity of meaning of the two sentences.

Sentences that are more complex should make the decision harder to make either 1) because more processing must be done before a meaning can be discovered or 2) because the decision procedure is more difficult due to the complexity. Either of the difficulties would be reflected in the reaction time measure because both the comprehension process and the decision process would have to be completed before any decision could be implemented by way of the buttons being pressed.

The materials consisted of a set of sentence pairs. The sentence pairs could consist of two idioms, one idiom and one non-idiom or two non-idioms. If idioms are more difficult to comprehend, two idioms should take longer to perform the decision task upon than one idiom, and one idiom should take longer than none at all. The degree of similarity was SAME (when the similarity of the pair was semantically similar but not physically identical), DIFFERENT (when the difference was semantic and physical), and IDENTICAL (when the similarity was physical, that is, when the sentences of the pair were identical).

A similar paradigm has been used by Posner and Mitchell (1967) to study depth of processing for simple classification tasks. They found differences between the time to decide that two stimuli were the same when they were either physically similar or similar in name, for example, AA versus Aa. They interpreted these differences as revealing different levels of processing required for a correct decision by these different types of stimuli.

The procedure used for this experiment with idioms was an extension of the paradigm used by Posner and Mitchell. The materials used in this idiom experiment were complete sentences which are considerably different from single letters, but the classifications were comparable: physical identity or name identity. Considering the Posner and Mitchell results, the IDENTICAL stimuli should be faster to make decisions about than SAME stimulus pairs. In order to make a correct decision on sentence pairs which were IDENTICAL, no extensive semantic decoding is required. Taking advantage of physical identity, it would be possible to respond correctly to these kinds of sentence pairs on the basis of physical similarity alone with little or no semantic decoding.

The sentence pairs with semantic similarity would require much greater processing before a correct decision of similarity could be made since their similarity is at a different level. These sentence pairs would necessitate a complex decoding and similarity matching operation before a correct decision could be made.

Posner and Mitchell's paradigm also included DIFFERENT decisions in which their stimuli had different names, A and B, for example. Their results indicated that there is no time difference between decisions of SAME and DIFFERENT. In the experiment with idioms, it was necessary to do the same kind of semantic decoding on the DIFFERENT sentence pairs as the SAME sentence pairs. Because of the lack of physical similarity, it was not possible to decide if the correct response was SAME or DIFFERENT until both sentences of these pairs were read, decoded, and matched on the basis of semantic similarity. It was, therefore, predicted that there should be no decision time difference between the SAME and DIFFERENT sentence pairs.

The necessity of semantic decoding for the SAME and DIFFERENT sentence pairs, and the method of random presentation of the sentence pairs was intended to force a more than superficial examination of all the sentence pairs. It would have been a bad strategy for subjects to try to make all decisions on the basis of physical similarity. Such a strategy would have produced correct responses to IDENTICAL sentence pairs, and even correct responses to DIFFERENT sentence pairs because of their physical dissimilarity, but not to SAME sentence pairs. In addition, such a strategy would not have necessitated the kind of semantic processing which would reveal any differences in the comprehension of idioms as compared to non-idioms, if such differences exist.

The predictions, then, were: that the IDENTICAL pairs should be easier due to the facilitation of the physical similarity (as found by Posner and Mitchell); that the SAME and DIFFERENT pairs should be harder than the IDENTICAL pairs because of the more extensive processing required; that there should be no difference in reaction time between SAME and DIFFERENT pairs (as found by Posner and Mitchell) because the same semantic processing is required for correct decisions on both kinds of pairs.

2.1. Pre-tests

2.1.1. Pre-test I

A group of idioms was obtained by way of observation of language use, and by reference to an idiom dictionary (Collins, 1958a & 1958b). This group consisted of 28 idioms, 14 pairs which the experimenter believed to be similar in meaning. Fourteen filler sentences were added to this group of 28 idioms, and these were given in random order in the form of a questionnaire to a group of 25 subjects. The subjects were asked to write a paraphrase for each sentence in the space provided on the questionnaire. The subjects were given ample time to complete this task.

Subjects were 25 University of Texas undergraduates who participated in the experiment as part of the requirement for an introductory course in psychology.

The subjects provided much more information than that required in the paraphrase task. For almost every idiom, several paraphrases were given. Both non-idiomatic paraphrases and paraphrases which were other idioms were given. This provided non-idiomatic meanings of the idioms and also more idioms with meanings similar to the idioms presented.

The idiom paraphrases were rated for similarity of meaning, both by the experimenter and by another person who was not involved in the design, materials or procedure of the experiment. On the basis of the independent similarity ratings by the experimenter, the other judge, and the further information supplied by the subjects themselves, it was possible to eliminate idioms that were unfamiliar or had variable meanings, and to obtain more idiom pairs with similar meanings. It was considered desirable to have as many idioms as possible in order to avoid materials effects idiosyncratic to only a small portion of idiomatic expressions.

2.1.2. Pre-test II

A second test of familiarity was devised. In this test, the subjects has the task of rating pairs of sentences as either 1) highly similar 2) somewhat similar 3) dissimilar or 4) not related in meaning. The pairs consisted of 17 idiom-idiom pairs found to be similar in pre-test I (using the extra information collected in the paraphrase task) and 15 pairs of filler sentences of varying degrees of similarity with respect to meaning.

The subjects were 20 University of Texas undergraduates who participated in the experiment as part of a requirement for a course in introductory psychology.

The subjects were instructed to place a number which represented the degree of similarity of meaning of the two sentences of the pair in the space following each sentence. They were instructed to place an X in the blank if they did not understand or were not familiar with either sentence of any pair.

According to the number ratings assigned by the subjects, an average rating was computed for each sentence pair. Twelve idiom pairs were selected which had an average similarity rating of 1.4 on the scale or four.

2.2. Materials and design

For each of the 12 idiom pairs selected from the second pre-test, an idiom-non-idiom pair with similar meaning was constructed by taking one idiom of the pair and matching it with its non-idiomatic meaning (as determined by taking the most frequent non-idiom paraphrase from the first pre-test). This non-idiomatic sentence was also matched with another which was similar in meaning (again selected from the paraphrase pre-test), yet non-idiomatic. This resulted in three pairs, one consisting of two idioms, one of one idiom and one non-idiom and the third of two non-idioms, which were all similar in meaning. An attempt was made to keep the length in letters of the sentences as close as possible.

For each pair from the idiom-idiom group, one idiom sentence was randomly selected and paired with another idiom of similar length which was dissimilar in meaning. This same idiom was paired with a non-idiomatic sentence which was dissimilar in meaning to form an idiom-non-idiom pair. One of the non-idiom sentences which was similar in meaning was paired with an unreleated sentence to form another non-idiom pair, this one dissimilar in meaning. This resulted in an idiom pair, an idiom-non-idiom pair and a non-idiom pair, all with dissimilar meanings. Again, an attempt was made to keep length constant across sentence pairs.

Each idiom used both in the similar and dissimilar meaning groups was paired with itself to form an idiom pair which was physically identical. The non-idiom counterpart of this was also created. See Table 1 for an example of the materials and the design.

Even though the attempt was made, it was not always possible to keep length constant within each cell; it was, however, possible to equate mean length across all cells. That is, the average of the idiom-idiom similar cells was equal to the average of the idiom-idiom different cells, etc., for all cells.

2.3. Procedure

The materials consisted of 140 sentence pairs: 12 pairs in each of those categories described above, plus 20 practice sentences and 24 filler sentences. The subjects were informed that they would see a series of sentence pairs, one pair at a time. Subjects were instructed to read each pair, decide if the two sentences in the pair were similar or different in meaning and to push one of the two buttons on the panel in front of them according to the decision they had made. Subjects were requested to read both sentences of a pair before making any decision, but to make their decisions as quickly as possible. Examples were given of similarity and difference of meaning, and questions answered, if necessary, to clarify this point. The subjects were informed that, in the course of the experiment, they would see the same sentence more than once, but they were cautioned that the pairing would be different each time and the decisions must be made on each pair.

The sentences were presented on a cathode ray tube display screen. Each pair was preceeded by the word READY which appeared in the upper left of the display screen one sec prior to the presentation of the sentence pair, and was removed immediately before the presentation of each pair. Each sentence pair was displayed beginning on the middle line of the screen and continuing consecutively to additional lines if length necessitated. The display of each pair was arranged such that no sentence was presented directly below the other since this type of

TABLE 1

EXAMPLE MATERIALS AND DESIGN FROM EXPERIMENT I

SAME materials (materials semantically similar)

Idiom-Idiom

They were putting John on.
They were pulling John's leg.

Idiom-Non-Idiom

They were putting John on.
They were teasing John.

Non-Idiom-Non-Idiom

They were teasing John.
They were kidding John.

DIFFERENT materials

Idiom-Idiom

They were putting John on.
He turned over a new leaf.

Idiom-Non-Idiom

They were putting John on.
We decided to do our work.

Non-Idiom-Non-Idiom

They were kidding John.
Ellen found two cups.

IDENTICAL materials

Idiom-Idiom

They were pulling John's leg.
They were pulling John's leg.

Non-Idiom-Non-Idiom

They were kidding John.
They were kidding John.

presentation would facilitate a template-matching strategy, and this was not desired. Each sentence pair was presented for five sec and removed at the end of that period, regardless of response. Each subject first saw 20 practice sentence pairs, which were designed to be of the same type as the materials used in the remainder of the experiment. A timer was started when the sentence pair appeared and was stopped when either of the two buttons was pressed, or when the sentences were removed from the screen, whichever came first. The presentation of the sentences, timing and recording of the button press responses were handled by a PDP 8/I computer.

2.3.1. Subjects

Subjects were 30 University of Texas at Austin undergraduates who participated in the experiment as part of their course requirement for an introductory course in psychology. Subjects were tested individually. After each subject had participated in the experiment, the purpose of the experiment was explained and questions answered.

2.4. Results and Discussion

Even though the idiom materials were extensively pre-tested, when dealing with idioms and with decision tasks, there is always some individual variation and some instances of incorrect responses. This variation did cause some incorrect responses. These error rates for all conditions are presented in Table 2. It was not appropriate to use the erroneous responses in the analysis. Therefore, it was necessary to discard the errors, causing missing entries in the raw data. Because of these missing entries, it was not feasible to use raw data in the analysis.

Instead, a mean was computed for each subject across the sentence pairs representing each cell, and the analysis performed on these means. Table 3 is a table of the mean of these subject means.

A quick glance at Table 3 shows that the pairs containing idioms were more difficult to make decisions about than the non-idiomatic materials. The pattern of results is especially clear in the DIFFERENT materials in which two idioms appear to cause more difficulty than one, and one seems to be harder than none. This pattern does not appear in the SAME materials because of the non-idiom pair. This cell mean is larger than the idiom-non-idiom when it should be smaller. This cell was a very difficult one for which to create materials because of the length, grammatical structure and semantic similarity constraints, and the resulting materials, while conforming to the stated criteria, were not always the most common way of stating the concept. The materials from this cell were, unfortunately, not pre-tested (since they did not involve idioms), so the cause of the difficulty is not known. But it seems likely that the difficulty is some kind of frequency effect since the materials from this cell were often of low frequency. All other idiom effects conform to the predictions.

One other effect is quite clear from a quick examination of the data: there is a difference between IDENTICAL, SAME and DIFFERENT sentence pairs. The IDENTICAL pairs were faster, as predicted, but there is also a consistent difference between SAME and DIFFERENT sentence pairs which was not predicted.

On first examination, the design appears to be a 3 x 3 with one missing cell. This is not entirely accurate since the cell is not missing because of oversight. Rather, that cell is nonexistent--there can be no such thing as a physically identical pair where one sentence is

TABLE 2

ERROR RATES FOR EXPERIMENT I

	IDIOM-IDIOM	IDIOM-NON	NON-NON
SAME	8%	9%	18%
DIFFERENT	3%	3%	2%
IDENTICAL	5%		2%

TABLE 3

RESULTS FROM EXPERIMENT I

Reaction times are shown in seconds

	IDIOM-IDIOM	IDIOM-NON	NON-NON
SAME	2.60	2.48	2.55
DIFFERENT	2.51	2.45	2.36
IDENTICAL	2.17		2.04

an idiom and the other is a synonymous non-idiom. Thus, the design is best considered to be two experiments with overlapping cells, one involving two sentence similarity types (SAME versus DIFFERENT) and all three idiom levels, and the other involving all three sentence similarity types but only two idiom levels (two idioms versus two non-idioms).

2.4.1. Analysis I

One analysis was performed on two levels of the similarity variable (SAME when materials were semantically similar versus DIFFERENT) by the three levels of the idiomaticity variable (idiom-idiom, idiom-non-idiom, non-idiom-non-idiom). Each of the 30 subjects responded to all conditions, so the design was a within subjects design with two variables, one with two levels and one with three.

The effect of the similarity type was significant, $F(1, 29)=7.54$, $p<.02$, with the SAME decisions taking longer. It has been found elsewhere (Posner & Mitchell, 1973) that difference-type decisions, are not made faster than similar-type decisions, even though it is theoretically possible. For these materials, this was not true. The SAME materials took longer to make decisions about. This could have been due to either 1) making a DIFFERENT decision on the basis of partial information from the second sentence of the pair or 2) some amount of extra time required to make a SAME decision. Subjects were instructed not to do 1), but that is no guarantee that they did not. However, there is some evidence from the second analysis that 2) is the more likely explanation; this will be discussed with the presentation of that analysis.

The effect of sentence type also was significant, $F(2,58)=6.73$, $p<.005$. The results were as predicted: it took longer to decide about two idioms than to decide about one non-idiom which, in turn, took longer than to decide about two non-idioms. Idioms, then, do seem to be more difficult to deal with psychologically than non-idiomatic expressions. The sentence type by similarity type interaction was also significant, F(2,58)=5.27, $\underline{p}<.01$, indicating that decision types are differentially affected by the type of material.

2.4.2. Analysis II

A second analysis of variance was performed using the data from three similarity types and two sentence types. The three similarity types were SAME (materials semantically similar), DIFFERENT, and IDENTICAL (when materials were physically identical). The two sentence types were idiom-idiom and non-idiom-non-idiom. The sentence type effect was significant in this analysis also F(1,29)=25.04, $\underline{p}<.0001$, again indicating that dealing with two idioms is more difficult than dealing with none. The effect of similarity type was also highly significant, F(2,58)=56.64, $\underline{p}<.0001$. The previous analysis indicated that SAME decisions took longer than DIFFERENT decisions when the materials were semantically similar. This analysis suggests that it is not the SAME decision per se that is difficult, but some component of the process which determines a semantic match, because the IDENTICAL decisions were much shorter than the SAME decisions. Both decisions required a response of SAME, but these decisions varied on the average by over 400 msec, indicating very different processing procedures. An attempt was made in the presentation of these IDENTICAL sentences to avoid a situation where 'template matching' could be used to make the match; the two sentences of a pair could not easily be matched on only physical criteria due to the arrangement of the pair on the screen. Hopefully, it was necessary for the subjects to at least read through both sentences, and the reaction times (over two sec, on the average) indicate that this was done. The indication is that physical identity makes a decision much easier than a semantic similarity, a result exactly in line with the Posner and Mitchell results. This is a very reasonable result considering that extensive semantic decoding as well as some kind of matching procedure is required in order to make such a decision. This process takes more time where there are fewer physical similarities since more extensive semantic decoding is then required.

The interaction of similarity type and sentence type was not significant, $F(2,58)=1.91$, $p>.10$) for this analysis.

The results of this experiment strongly indicate that idiom comprehension is not only different from the normal course of comprehension (as indicated by the results of Bobrow and Bell), but is also more complex and/or difficult (or at least takes longer). The experiment also indicates that the type of decision made about the materials differs both according to similarity of the sentence pair and according to the type of materials on which the decision is made.

CHAPTER III

THE SECOND IDIOM EXPERIMENT--THE AMBIGUITY VARIABLE

The first question about idioms has an answer: Idioms do seem to be handled differently from non-idioms in comprehension tasks. The other questions that were raised earlier now become relevant. What effects do the different types of idioms have on the comprehension process? Since idioms are either ambiguous or ungrammatical, is this a clue as to the nature of their processing? That is, might it be that ambiguous idioms are processed like other ambiguities and ungrammatical idioms like other ungrammatical sentences?

Consider the variable of ambiguity. As previously discussed (Section 1.1) ambiguities make comprehension more difficult in many cases. In the materials from Experiment I there was no distinction made between ambiguous (grammatical) and unambiguous (ungrammatical) idioms. In all cases, the control non-idiomatic sentences were unambiguous. It is conceivable that this is important: perhaps the idiom effect there was really only an ambiguity effect, obtained because some of the idiomatic materials were ambiguous. Perhaps there is no such thing as an 'idiom effect' per se, but only an ambiguity effect, an effect which has been found by many experimenters.

Intuitively, this does not seem to be the case. As many investigators have pointed out (MacKay, 1966; Garrett, 1970), ambiguities go largely unnoticed in the course of sentence comprehension. This is not the case with idioms; they are readily noticed.

One kind of ambiguity which does get noticed is ambiguity which causes garden-pathing, the unexpected and necessary recomputation of the other meaning of an ambiguity, such as in Sentence (11). This recomputation

(11). The fat grocer weighed 300 grapes.

is necessitated when it becomes apparent, through context, that the inappropriate reading has been assigned to an ambiguity. In sentence (11) the recomputation is usually necessary at the word *grapes* because it does not fit with the meaning usually computed for the previous part of the sentence. Perhaps some similar process occurs with idioms. With the ambiguous idioms, one might get the literal reading, notice it did not fit, i.e., experience garden-pathing, and recompute. For the unambiguous idioms, one would try to get a reading, notice the anomaly, and recompute. This seems like a simple but reasonable model, but the analogy is not perfect. Garden-pathing occurs because of the context, and ambiguous idioms are interpreted as idioms even in the absence of context, an Bobrow and Bell have demonstrated.

Even though idioms and non-idiomatic ambiguities seem, intuitively, to be different kinds of phenomena, intuition is not always a reliable guide concerning language processing. It is possible that idiom comprehension proceeds in the same way as ambiguity comprehension because they are parts of the same phenomenon.

If this is the case, then there is, again, the question: What kind of ambiguities are idioms like? As previously discussed, Bobrow and Bell (1973) investigated this question. They used both lexical and surface structure non-idiomatic ambiguities as materials in their 'set procedure', but found no significant effect for either of these types of materials in facilitating an idiomatic interpretation. Their results are not completely clear, however, since they did not have a control set of unambiguous literal materials to which to compare the ambiguous sets. This leaves unclear whether ambiguity might not be a factor in idiom comprehension (because the baseline response rate was the same as the response rate after both ambiguous sets) or because any kind of ambiguity facilitates an idiomatic reading (since the structural and lexical set groups were the same). This is certainly an important question.

Idioms do not seem to be like surface structure ambiguities. Their ambiguity does not depend on the grouping of the words, nor can it be determined by intonation. Ambiguous idioms do seem to be something like lexical ambiguities in that they seem to focus on an item, but they are unlike lexical ambiguities in that they involve interpretation of several items. But the interpretation of several items in an idiom is often synonymous with a single word. Ambiguous idioms seem to be somewhat like deep structure ambiguities, but instead of two different deep structures in which the two derivations converge, idioms have one full structure and one aberrant one. The aberrant structure is not related semantically to the full structure in that the meaning of the words is entirely different from the meaning of the idiomatic phrase; the two structures share no semantic information.

Idioms seem to have similarities to both lexical and deep structure ambiguities. Very little is known about the processing of deep structure ambiguities, but the difficulties which are caused by lexical ambiguities are more familiar. For this reason, and because of the length constraint, lexical ambiguities were chosen as the non-idiomatic ambiguities for this experiment.

3.1. Design and materials

An experiment was designed to evaluate the role of ambiguity in the comprehension of idiomatic expressions. The same type of design and procedure was used in this experiment as was used in Experiment I, with ambiguity as an additional variable within this design. All of the similarity types from Experiment I were present, but the ambiguity variable increased the design to twice the size of the design in Experiment I. See Figure 1 for a sketch of the design.

The materials from the first experiment included the following types of similarity judgments: SAME (when the sentences were semantically similar), DIFFERENT (when the sentences were semantically and physically dissimilar), and IDENTICAL (when the sentences were physically identical). The sentence types were: idiom-idiom, idiom-non-idiom, and non-idiom-non-idiom.

Creating materials with the above characteristics and maintaining ambiguity as a variable was extremely difficult in many cases, and completely impossible in some. To follow all the criteria used to produce materials in Experiment I and add the ambiguity variable as well would necessitate the creation of an idiom pair, an idiom-non-idiom pair and a non-idiom pair all members of which would have the same meaning and be ambiguous. Perhaps one such example exists, but that would not be sufficient materials for an experiment. It is possible (with some difficulty) to create the idiom-idiom pairs, but not to find non-idiom sentences to match with these which are both semantically similar and ambiguous. Consequently, the decision was made to forego the ambiguity variable for those two impossible cells and to retain the semantic similarity criterion, preserving as much similarity to the first experiment as possible. It was possible to retain ambiguity within the remainder of the design, creating ambiguous idiom pairs and ambiguous non-idiom pairs with different meanings, ambiguous idiom pairs

FIGURE 1

A SKETCH OF THE DESIGN USED IN EXPERIMENT II

AMBIGUOUS

	IDIOM-IDIOM	IDIOM-NON-IDIOM	NON-IDIOM-NON-IDIOM
SAME		**	**
DIFFERENT	*		*
IDENTICAL	*		*

UNAMBIGUOUS

	IDIOM-IDIOM	IDIOM-NON-IDIOM	NON-IDIOM-NON-IDIOM
SAME			
DIFFERENT	*		*
IDENTICAL	*		*

* Cells analyzed

** Impossible by original criteria

with similar meanings, and physically identical pairs with, of course, identical meanings. The ambiguous non-idiomatic materials were all lexical ambiguities. In selecting lexical ambiguities, an attempt was made to select ambiguities which were unbiased, that is, whose meanings are close to equiprobable. Lexical ambiguities were chosen in order to test the hypothesis that idioms are processed like lexical ambiguities. If the reaction times were not significantly different or showed the same pattern of response times, it could be said that idioms are processed like lexical ambiguities. If neither of these possibilities occur, then some alternative explanation must be chosen.

Of course, it was possible to create materials which conformed to each of the categories which were unambiguous. See Table 4 for an example of the type of materials used in this experiment.

The lack of strict adherence to the ambiguity variable raises serious problems for any analysis. It would not be correct to analyze this design in the same way as that of Experiment I because two of the cells do not conform to the ambiguity variable. It was possible to resolve this problem in what seemed to be a satisfactory way. If the same pattern of results was obtained in this experiment as in the first experiment, it would be possible to look selectively at some of the cells and see the entire pattern of the results. In Experiment I, the pattern of results is clearly revealed in the reaction times representing the idiom-idiom and non-idiom-non-idiom DIFFERENT and IDENTICAL cells. If these effects also exist in this design, these cells should reveal them. The decision was made to analyze only the two similarity types, DIFFERENT and IDENTICAL and the two sentence types, idiom pairs and non-idiom pairs for each ambiguity condition, ambiguous and unambiguous. This makes a total of eight cells in

TABLE 4

AN EXAMPLE OF MATERIALS USED IN EXPERIMENT II

AMBIGUOUS	UNAMBIGUOUS
SAME materials	SAME materials
Idiom-Idiom	Idiom-Idiom
He was afraid to break the ice.	My refrigerator is out of whack.
He was afraid to take the first step.	My refrigerator is on the blink.
Idiom-Non-Idiom	Idiom-Non-Idiom
He was afraid to break the ice.	My refrigerator is out of whack.
He was afraid to begin.	My refrigerator is not working.
Non-Idiom-Non-Idiom	Non-Idiom-Non-Idiom
He was afraid to begin.	My refrigerator is not working.
He was afraid to start.	My refrigerator is not running.
DIFFERENT materials	DIFFERENT materials
*Idiom-Idiom	*Idiom-Idiom
He was afraid to break the ice.	My refrigerator is out of whack.
He was really out on a limb.	My aunt hauled him over the coals.
Idiom-Non-Idiom	Idiom-Non-Idiom
He was afraid to break the ice.	My refrigerator is out of whack.
He found a bat in the attic.	My father criticized the election.
*Non-Idiom-Non-Idiom	*Non-Idiom-Non-Idiom
He found a bat in the attic.	My father critized the election.
He had a flat in downtown London.	My sister likes her history professor.
IDENTICAL materials	IDENTICAL materials
*Idiom-Idiom	*Idiom-Idiom
He was afraid to break the ice.	My refrigerator is out of whack.
He was afraid to break the ice.	My refrigerator is out of whack.
*Non-Idiom-Non-Idiom	*Non-Idiom-Non-Idiom
He had a flat in downtown London.	My sister likes her history professor.
He had a flat in downtown London.	My sister likes her history professor.

* Cells used in the analysis.

a 2 x 2 x 2 factorial design. It was decided that the other cells, an additional eight, would be included in the materials but not analyzed so that subjects would have to make the same kinds of decisions in this experimental design as the subjects in the first experiment did.

3.1.1. Pre-test

After materials conforming to the above criteria were obtained, a procedure to determine the acceptability and familiarity of these materials was constructed. This test was the same sort as that used for Pre-test II of Experiment I, a similarity rating task. Subjects were asked to rate sentence pairs as either 1) highly similar, 2) slightly similar, 3) dissimilar, or 4) not related in meaning. Sentence pairs were from all of the categories described above, except the categories which involved physically identical materials, a total of 168 sentence pairs.

Eleven sentence pairs from each category were chosen to be used in the experiment. The ratings were used to determine the most familiar idioms and those with the most consistent meanings. The idioms were chosen on the basis of similarity judgments on the idiom pairs, even though these pairs would not be analyzed, because if the idiom pair upon which the materials for the remainder of the group is based is familiar, then the remainder of the group will be familiar. The ratings are shown in Table 5. Note that there is no rating for the IDENTICAL group since these were not pre-tested.

3.2. Procedure

The procedure was exactly the same as that used in Experiment I. The subjects were informed that they would see a series of pairs of sentences, one pair at a time. Each subject was instructed to read both sentences of each pair, decide if they were similar or different in meaning, and to press one of two buttons on the panel according to the decision. Each subject was informed that in order to make a correct decision, it would be necessary each time to read both sentences of a pair. Subjects were requested to make their decisions as quickly and as accurately as possible. They were given examples of similarity and difference of meaning, and any questions were answered in order to clarify this point.

The sentence pairs were presented on a cathode ray tube display screen. Each pair was preceeded by the word READY which appeared in the upper left of the screen one sec prior to the presentation of the sentence pair. The sentence pairs were displayed on the screen beginning on the middle horizontal line of the display tube, and continued to following lines if length necessitated. The sentences were arranged such that no one sentence of the pair appeared directly above the other. The sentence pairs remained on the screen for five sec before being removed. The sentences were removed from the screen after the five-sec interval, regardless of response. Each subject first saw 20 practice pairs of the same sort as the pairs in the experiment. A timer was started when the sentence pair was presented and stopped either when a button was pressed or when the sentence pair was removed from the screen, whichever came first. The presentation of the sentences, timing and recording of responses was handled by a PDP 8/I computer.

3.2.1. Subjects

Subjects were 40 University of Texas at Austin undergraduates

TABLE 5

RATINGS ON PRE-TEST FOR EXPERIMENT II

The range of the scale was one to five

AMBIGUOUS

	IDIOM-IDIOM	NON-IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
SAME	1.8	1.6	1.6
DIFFERENT	4.6		4.7

UNAMBIGUOUS

	IDIOM-IDIOM	NON-IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
SAME	1.7	1.6	1.6
DIFFERENT	4.7		4.7

who participated in the experiment as part of the requirement for an introductory course in psychology. Subjects were tested two at a time or individually, depending on the availability of subjects. After participation in the experiment each subject was given written explanation of the purpose of the experiment, and any questions about the nature of the experiment were answered.

3.3. Results and Discussion

The treatment of the results from this experiment was very much like that used for the data from the first experiment. Again, some provision had to be made for errors. The same procedure was carried out in this experiment as in Experiment I; the erroneous responses were discarded, and a mean for each subject for each cell was computed using the correct responses. The analysis was done on these means rather than on raw data. The error rates for this experiment are shown in Table 6. Note the difference in error rates for the ambiguous and unambiguous IDENTICAL cells. The ambiguous materials produce much greater error rates. This is interesting because the decision that two identical sentences mean the same thing should be a simple one and should produce few errors. It is possible to make the decision that two identical sentences are different in meaning for ambiguities if both meanings of the ambiguity are found and assigned to the sentences of the pair. But to make this decision would seem to entail much more work for the comprehension mechanism. Yet something like this seems to be going on since the error rates for the unambiguous cells are not nearly as large.

The means of the subjects' mean reaction times for each cell are shown in Table 7. As in Experiment I, the responses to IDENTICAL

TABLE 6

ERROR RATES FROM EXPERIMENT II

AMBIGUOUS

	IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
DIFFERENT	2.9%	1.1%
IDENTICAL	5.9%	4.3%

UNAMBIGUOUS

	IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
DIFFERENT	1.6%	1.1%
IDENTICAL	.9%	1.6%

TABLE 7

RESULTS FROM EXPERIMENT II

Reaction times given in seconds.

AMBIGUOUS

	IDIOM-IDIOM	IDIOM-NON-IDIOM	NON-IDIOM-NON-IDIOM
SAME	*	*	*
DIFFERENT	2.85	*	2.96
IDENTICAL	2.43		2.74

UNAMBIGUOUS

	IDIOM-IDIOM	IDIOM-NON-IDIOM	NON-IDIOM-NON-IDIOM
SAME	*	*	*
DIFFERENT	3.01	*	2.83
IDENTICAL	2.38		2.46

* Cells not analyzed, but included.

materials were faster than those to DIFFERENT materials for all cases. However, the longer decision times for idiomatic materials which were found in Experiment I were not found in this experiment. There is no especially large ambiguity difference, but reaction times to the ambiguous non-idiomatic (lexically ambiguous materials) do seem to be considerably longer than to their unambiguous counterparts.

Since each subject received each level of each condition, the design was a within-subjects design. There were three factorial variables in the design, each having two levels: similarity type (IDENTICAL vs. DIFFERENT); sentence type (two idioms vs. two non-idioms); and ambiguity (ambiguous vs. unambiguous materials).

The analysis of variance revealed that ambiguity was a significant variable, $F(1,39)=14.49$, $\underline{P}<.005$, indicating that ambiguous materials, on the average, do take longer in this kind of comprehension task. The effect for type of similarity of materials was also significant, $F(1,39)=118.10$, $\underline{p}<.0001$. As in the first experiment, DIFFERENT materials took longer to make decisions on than IDENTICAL materials. The sentence type effect was also significant, $F(1,39)=13.36$, $\underline{p}<.005$, but in this experiment the non-idiomatic materials took longer than the idiomatic, just the opposite of what was found in Experiment I.

Examination of the ambiguity by sentence type interaction is somewhat revealing. This interaction was also significant $F(1,39)=41.91$, $\underline{p}<.001$, which indicates that the ambiguous non-idioms were processed differently from the unambiguous non-idioms. This is the same type of ambiguity effect found by other investigators which indicates that sentences with ambiguities are harder to deal with in some kinds of comprehension tasks. But idioms seem to be just the opposite. A t-test performed on the idiom-idiom DIFFERENT cells showed that the unambiguous idioms took longer to make decisions about than the ambiguous idioms, $t(39) = -3.14$, $p < .005$, whereas a similar test performed on the non-idiom DIFFERENT cells indicates that the ambiguous non-idioms took longer than the unambiguous $t(39) = 2.59$, $p < .02$. The obtained ambiguity effects for non-idioms not only give some confidence in the manipulation performed in the experiment, but also give some confidence in the task. They indicate that the comprehension task used for these experiments with idioms is sensitive to the same variables of comprehension as the comprehension tasks used by other investigators.

The interaction of the ambiguity and sentence type variables and the opposite effects of ambiguity found in the two t-tests indicate that ambiguity alone is not a sufficient explanation for the difficulty in comprehending idioms: it suggests that ambiguity differently affects the comprehension of idioms and non-idioms. If idiomaticity and ambiguity were facets of the same phenomenon, then the reaction times should reveal the same pattern for both idioms and ambiguities. This was not found to be the case, instead, it was found that idioms (whether ambiguous or not) do not take as long to deal with in this task as do non-idiomatic ambiguities, and even that unambiguous idioms take longer to make decisions about than ambiguous idioms. There is, then, the indication that the difficulty in comprehension of idioms is not the same thing as the difficulties found in comprehending ambiguous expressions which are not idiomatic; that there could be two separate phenomena, one of idiomaticity and one of ambiguity.

One other possibility that cannot be ruled out is that of a materials effect. There is, of course, always the possibility of a random materials effect, but when working with ambiguities, there is an additional possibility of some non-random bias in the materials.

Two other interactions were significant: the ambiguity by similarity type interaction, F(1,39)=25.84, $p<.005$; and the sentence type by similarity type interaction, F(1,39)=27.32, $p<.005$. These results indicate that decisions about the similarity of materials (IDENTICAL or DIFFERENT) are differently affected by both the presence of ambiguity and the presence of idiomatic materials. Decisions take longer for IDENTICAL materials when they are ambiguous but take a shorter time when they are DIFFERENT. The idioms took a shorter amount of time than non-idioms for IDENTICAL materials but longer for DIFFERENT materials. This seems to indicate the lack of a clear-cut difference between decisions on IDENTICAL materials and DIFFERENT materials. The presence of these two interactions is difficult to account for and prohibits the formulation of any coherent explanation of the nature of the idiom comprehension process. Although some of the results indicate the separability of idiomaticity and ambiguity, these two interactions and the sentence-type main effect, which indicated that idioms were easier than non-idioms, cast doubt on the results of the entire experiment. The interactions might be expected on the basis of Experiment I, since the interaction of similarity type and sentence type was significant in that experiment. But the addition of the ambiguity variable is a further complication. The sentence type effect is just the opposite of the one found in Experiment I, a very puzzling result.

Since the only controlled difference in materials and procedure between Experiments I and II was the presence of ambiguity, this was considered to be the first place to seek an explanation of these otherwise unexplicable results.

3.4. Bias

Even though the variable of ambiguity was well-controlled in the experiment, there are other factors within that variable. One which was not considered in the design was that of bias. In choosing the ambiguities for the non-idiomatic ambiguity conditions an attempt was made to choose 'unbiased' ambiguities. Such ambiguities are those items where the meanings, without some context, are close to equiprobable. That is, if a group of people was asked to define such a word, roughly one-half of the group would give one of its meanings and the other half would give the other meaning. For example, if the word TANK was given, about half the people asked for a definition would define it as "a storage container of some sort" and the other half as "an armoured vehicle".

The role of bias in the comprehension of ambiguous sentences has been investigated (Foss, Bever & Silver, 1968; Cairns, 1970, 1973). These studies do indicate that bias does have some effect on comprehension, but the tasks used were designed to necessitate the discovery of the less probable meaning of a biased ambiguity. The task used in Experiment II is not directly comparable; if anything, it is designed to do just the opposite, necessitate the discovery of the most frequent meaning. Therefore very little can be learned from these experiments other than the possibility that bias may have some effect on comprehension processing.

As previously mentioned, an attempt was made to select unbiased ambiguities. Reprocessing should not have been necessary. In the cases where a decision was made on IDENTICAL materials, the two sentences were physically identical and in the cases where the materials were physically different, reprocessing could have occurred, but should not have been necessary for a correct response. However, it is conceivable that bias did play a role in Experiment II. In the comprehension task used in these idiom experiments, bias could have caused the decisions about ambiguous materials to be easier or harder, depending on the bias of those materials.

The bias of the ambiguous idioms is unknown but, intuitively, it would seem as though the idiom readings would be more likely, making these biased ambiguities. Chafe (1968) has pointed out that this seems generally the case for ambiguous idioms. It is certainly the case for the unambiguous idioms; there is no parallel literal grammatical reading for such expressions. There is, theoretically, no ambiguity, and no sensible reading other than the idiomatic one. If Chafe's assertion is correct, then the idioms in the experiment would have been biased and the non-idioms unbiased. In that case, another variable, that of bias, would have been operating within the design and the comparisons made would have been confounded. However, assertions made about linguistic phenomena, even by linguists, do not necessitate the reality of those phenomena. Therefore, an experiment was designed to determine the bias of the materials used in Experiment II.

3.4.1. Materials and Design

The materials were taken from the relevant cells of Experiment II; those which contained either idioms, non-idiomatic ambiguities, or their unambiguous, non-idiomatic controls, eight cells in all. These sentence pairs were separated into single sentences, and 42 filler sentences, selected from other cells of the design, were added, making a total of 130 sentences. The categories were 1) sentences with lexical ambiguities; 2) sentences with ambiguous idioms; 3) sentences with unambiguous idioms and 4) sentences with unambiguous non-idiomatic words and phrases. There were 22 sentences of each of these four types, plus the filler sentences. These were separated into two sets of materials, one with 60 sentences and the other with 70 sentences. The division was made by dividing the ambiguous idioms randomly into two groups, along with those sentences that were length controls for them in the experiment. This would make a division of all four types of sentences used. The filler sentences, being taken from the experiment also, were divided according to this same pattern. This separation into sets was made due to the length of the task.

3.4.2 Procedure

The task used was a paraphrase task. Subjects were instructed to "write another sentence that means the same thing as the sentence given" for each sentence. Adequate space to write a paraphrase was provided. Subjects were instructed that, if they did not understand a sentence, they should put an X in the blank which followed each sentence.

After subjects were finished with the paraphrase task, a rating task was done by the subjects on their own papers, a procedure suggested by Cairns (1970). For each sentence which contained a possible ambiguity, either idiomatic or literal, the two possible meanings were read aloud by the experimenter. Subjects were asked to indicate by marking either (a) or (b) in the blank after each sentence which of the meanings they used in their paraphrases, or, if they used a third meaning to indicate so by marking (c) in the blank provided. The procedure was identical for each of the two groups.

3.4.2.1. Subjects

Subjects were 110 University of Texas undergraduates, 55 in each group, who participated in the experiment as part of a course requirement for a course in introductory psychology. After participating in the experiment, each subject was given a written explanation of the purpose of the experiment, and any questions about the experiment were answered.

3.4.3. Results and Discussion

Bias was determined for each of the 22 sentences in the following categories: 1) lexical ambiguities; 74%; 2) idiomatic ambiguities, 89%; 3) idioms which (theoretically) were not ambiguous (that is, have no parallel literal interpretation), 98%.

Chi-square analysis were performed on these frequency data. In order to determine whether each group had a significant bias, the more frequent response was compared to the less frequent. If the items in a group were not significantly biased, this frequency difference would not obtain significance. The scores for each of the groups did obtain significance: 1) lexical ambiguities ($\chi^2=264.2$, $p<.01$), 2) ambiguous idioms ($\chi^2=660$, $p<.01$), and 3) unambiguous idioms ($\chi^2=1028.9$, $p<.01$). This indicates that all sentence types were biased. The idioms were biased, as predicted, with the idiomatic reading being the more frequent for every item. However, the ambiguous lexical items were also biased, indicating that the readings of the materials of this type used in the second experiment, as a group, were not equiprobable, although some items within the group were.

A chi-square analysis was performed to determine if the extent of the biases between these groups was significantly different. If they were, then the possibility exists that the difference in bias could have affected performance in the comprehension task from Experiment II in the ways previously mentioned. The analysis was performed by comparing the frequency of the more frequent response for one group versus the more frequent response for the other and the frequency of the less frequent in the same manner. It was found that the lexical ambiguities were significantly less biased than the ambiguous idioms which were, in turn, less biased than the unambiguous idioms ($\chi^2=68.5$, $p<.01$). This reveals that there were, indeed, significant differences in the degree of bias for these three groups, a difference that could have affected performance in the comprehension task used in the second experiment.

In order to determine the effect of bias on reaction time in the comprehension task from Experiment II, correlations between the bias scores and the reaction times were calculated. In order to take into account the variable of sentence length, simple (product-moment) correlation coefficients for sentence length and reaction time and for bias and reaction time were determined, and also a partial correlation between bias and reaction time which would take sentence length into account, as well as a partial correlation between sentence length and reaction time, which would correct for bias. These correlations were done for each of the three sentence types individually by similarity type (N=11 for each). The same procedure was followed to obtain further correlations taken across the three similarity types (N=33). This produced eight partial correlations of length with reaction time and eight of reaction time with bias. If bias does facilitate reaction time in the comprehension task, then the bias and reaction time variables should be negatively correlated; as bias increases, reaction time should decrease. The correlations are shown in Table 8. Only one of the partial correlations of bias and reaction time

TABLE 8

CORRELATIONS WITH BIAS

PARTIAL CORRELATIONS OF BIAS WITH REACTION TIME (CORRECTED FOR LENGTH)

	AMBIGUOUS		UNAMBIGUOUS	AVERAGE OVER GROUP
	IDIOM-IDIOM	NON-IDIOM-NON-IDIOM	IDIOM-IDIOM	
IDENTICAL	.35	-.44*	-.13	-.17
DIFFERENT	.05	.15	.41	.16

PARTIAL CORRELATIONS OF REACTION TIME WITH LENGTH (CORRECTED FOR BIAS)

	AMBIGUOUS		UNAMBIGUOUS	AVERAGE OVER GROUP
	IDIOM-IDIOM	NON-IDIOM-NON-IDIOM	IDIOM-IDIOM	
IDENTICAL	.05	.86**	.81**	.66***
DIFFERENT	.85*	.31	-.16	.27
	N=11	N=11	N=11	N=33

* $p<.06$

** $p<.005$

*** $p<.01$

approached significance ($r=-.439$, $p<.05$), this being for ambiguous lexical items, IDENTICAL materials. None of the other IDENTICAL materials' decisions approached significance, including the IDENTICAL materials taken as a group. None of the correlations between bias and reaction time approached significance in the right direction for any DIFFERENT materials' decisions. The variable of bias, although present, does not seem to have had any effect on the comprehension process.

If length significantly affects reaction time, these two variables should be positively correlated. The results of the correlations of length with reaction time were as puzzling as those for the correlations of reaction time with bias. The correlations with length when the materials were identical was significant ($r=.66$, $p<.01$) as a group, but for only two of the sentence types within this group were the correlations significant; the ambiguous lexical items ($r=.86$, $p<.005$) and the unambiguous idioms ($r=.81$, $p<.005$), while the ambiguous idioms showed absolutely no effect of length ($r=.05$, $p>.25$), precluding any explanation of a unified effect here. The DIFFERENT decisions seem almost opposite, the ambiguous idiom group was the only significant one ($r=.85$, $p<.005$), with neither other group, nor the entire DIFFERENT items as a group, showing any effect at all. Even sentence length does not seem to have any reliable effect in this experiment.

Several facts about the materials should be kept in mind. First, the sample sizes (11 of each type) were rather small, making significance especially difficult to obtain. Second, the honogeneity of response to the idioms was impressive, and this made for even greater difficulty in obtaining significant correlations. However, these only compounded, not created, the entire problem.

The question of the appropriateness of the type of ambiguity used arises again. Perhaps the choice of lexical ambiguities was the wrong one; perhaps deep structure ambiguities would have been a more revealing source of ambiguity. Perhaps the comparison of idiomatic and lexical ambiguities was a comparison of apples and oranges, resulting in a sticky mess.

This investigation was, at least, valuable in giving support to the claim that the idiomatic reading is the more frequent for those sequences that have two possible readings, one idiomatic and one literal. It also supports the difference between ambiguous and unambiguous idioms, which did differ in frequency of idiom responses. The bias survey indicated that for the unambiguous idioms, very few people could get a meaning which was not idiomatic.

Perhaps this is the next path to explore in search of an explanation. Perhaps the difference between ambiguous and unambiguous idioms is something besides the presence of ambiguity. Rather than considering idioms as ambiguous or unambiguous, one could just as correctly consider idioms as grammatical or ungrammatical. And this may be a more revealing way to consider them.

CHAPTER IV

CONSIDERING THE VARIABLE OF GRAMMATICALITY

The bias survey indicated that the two kinds of idioms, ambiguous and unambiguous, did differ in average bias ratings. The unambiguous idioms were consistently (98%) given idiomatic readings; the ambiguous idioms were also given idiomatic readings, but not as consistently (89%). There are two more pieces of evidence that indicate that these two types of idioms differ. First, the patterns of the correlations of reaction time with bias did differ; the patterns were not clearly interpretable for either group, but they were different patterns. The difference of patterns does suggest the separability of these two types of idioms in comprehension tasks. The second suggestion of a difference can be seen in Experiment II. By referring back to Table 7, it can be seen that the average reaction times indicate that the ambiguous idioms are not as difficult as the unambiguous ones. Neither Experiment II nor the bias survey is clearly interpretable, but these two suggestions of a difference merit further examination.

Recall that ambiguity or the lack of ambiguity of idioms coincides perfectly with another factor: grammaticality. Perhaps it is this factor which is the important one. Since studying idiomaticity within the framework of ambiguity has revealed little, perhaps it would be more profitable (as well as more accurate) to consider idioms within the framework of grammaticality and of the research that has been done on that variable of comprehension.

Unambiguous idioms are also grammatically deviant whereas ambiguous idioms are well-formed. Perhaps it is this grammatical deviance which creates the observed difference in the comprehension of idioms. It has been shown in previous experiments (Downey & Hakes, 1968; Danks, 1970) that grammatical deviance does seem to play a role in the comprehension of non-idiomatic sentences, but the role of grammaticality in the investigation of the comprehension of idioms has been totally ignored in the few investigations done previously.

An experiment was designed to test the role of grammatical deviance in the comprehension of idiomatic and non-idiomatic sentences. The experiment was designed to include idioms, both ambiguous (grammatical) and unambiguous (ungrammatical), and non-idioms, both grammatical and ungrammatical. The purpose of the experiment was to determine how idioms of both types compare to grammatical and ungrammatical non-idioms in a comprehension task.

The type of comprehension task used in the previous two experiments was obviously not suitable since it is not possible to judge the semantic similarity of two sentences which are ungrammatical, that is, which have no interpretable meaning. It was, therefore, necessary to change the comprehension task. It was decided that an acceptability judgment task would be suitable. In this kind of task, it is necessary to read a single sentence, decide if it is acceptable or not and to press a button according to the decision. This task was chosen not only because it was workable, but also because it preserved many factors from the similarity judgment task: the visual presentation, the YES-NO decision types, the button press responses. The results from these two tasks are not directly comparable, first because the presentation of two sentences takes more reading time than only one sentence, and second, because a similarity judgment task is a different kind of task than a judgment of acceptability. However, it seems reasonable to suppose that these two tasks are sensitive to the same kinds of processes of comprehension in that both entail a judgment about the meaning of sentences and in order to make such a judgment the meaning of the sentence must be decoded.

4.1. Design and Materials

By searching the materials collected in preparation for the previous experiment, a large group of idioms was obtained. This group was divided into two groups, one of grammatical idioms and one of ungrammatical idioms. From this collection of idioms, a group of 30 of each grammaticality type was selected which were believed to both adhere strictly to the grammaticality/ungrammaticality distinction and be familiar to most people. For each idiom a non-idiomatic, grammatical counterpart was constructed, using the same grammatical structure and words of the same frequency (Thorndike & Lorge, 1944). In addition, an ungrammatical counterpart for each idiom was constructed using the same grammatical structure and words of the same frequency, but in each sentence substituting one word which would violate the selectional restrictions in that structure, making the sentence anomalous. For example, for an idiom such as HE WAS DOWN IN THE DUMPS, a grammatical sentence with the same grammatical construction, word frequencies and word lengths was created, such as HE WAS DOWN BY THE PIER. An ungrammatical sentence with the same grammatical construction, and words of the same lengths and frequencies was created by substituting words which were inappropriate for the grammatical construction in the sentence frame to make such a sentence as HE WAS OVER IN THE PANG.

There was, thus, a total of six groups of sentences: two idiom groups, one with grammatical and one with ungrammatical idioms; two non-idiom grammatical groups, created as length, word frequency and structure

TABLE 9

EXAMPLES OF MATERIALS USED IN EXPERIMENT III

TYPE I	
NON-IDIOM GRAMMATICAL:	He took the cup out of the box.
IDIOM GRAMMATICAL:	He let the cat out of the bag.
NON-IDIOM UNGRAMMATICAL:	He let the dog out of the hope.
TYPE II	
NON-IDIOM GRAMMATICAL:	She had to wash her hair.
IDIOM UNGRAMMATICAL:	She had to eat her words.
NON-IDIOM UNGRAMMATICAL:	She had to sell her heat.

controls for each group of idioms; and two groups of ungrammatical non-idioms created by the same process and for the same reason as the grammatical non-idiom groups. This made a total of 180 sentences in all.

There were two variables: type of materials, Type I and Type II, defined by the two kinds of idioms. Type I consisted of the ambiguous (grammatical) idioms, plus the grammatical and ungrammatical non-idiomatic controls created for the ambiguous idioms. Type II consisted of the unambiguous (ungrammatical) idioms and their two kinds of control sentences. The other variable was grammaticality: grammatical non-idioms, idioms, and ungrammatical non-idioms. An example of the design and materials for this experiment can be seen in Table 9.

4.2. Pre-test

A pre-test was given in the form of an acceptability rating questionnaire. The materials used were the 180 sentences previously described. Due to the length of the task, the materials were divided in half by taking the first 15 sentences of each group to make one list, and the last 15 to make the other. The sentences were randomized within these two lists, making two lists of 90 sentences each. The testing was conducted in two parts corresponding to the two lists. Subjects were instructed to read each of the 90 sentences and to decide if each was acceptable or not. Subjects were told that acceptability could be defined in terms of meaningfulness, familiarity, grammaticality, or some combination of these, but that acceptability if often closely related to intelligibility. They were instructed to rate the sentences either as 1) fully acceptable, 2) not as acceptable as other sentences, 3) unacceptable or 4) completely unacceptable. Examples

TABLE 10

PRE-TEST RATINGS FROM EXPERIMENT III

The scale was from zero to three.

	GRAMMATICAL NON-IDIOM	IDIOM	UNGRAMMATICAL NON-IDIOMS
TYPE I	.27	.31	2.45
TYPE II	.29	.32	2.44

were given of what would be generally considered acceptable and unacceptable sentences, and any questions necessary for clarification of this concept were answered.

The subjects were 60 University of Texas undergraduates, 30 in each group, who participated in the experiment as part of the requirement for a course in introductory psychology.

Acceptability ratings were tabulated on the basis of the ratings given by the subjects and a mean rating was computed for each of the experimental sentences. From the total of 30 idioms of each kind, a group of 16 grammatical and 16 ungrammatical idioms was selected, along with the appropriate number of non-idiomatic control sentences, a total of 96 sentences. The average ratings for each type are shown in Table 10. A rating of zero would indicate complete acceptability, and a rating of three would indicate complete unacceptability. The important consideration is that the sentences which are acceptable all have very similar and very good acceptability ratings whereas the unacceptable sentences have very poor acceptability ratings.

4.3. Procedure

Subjects were informed that they would see a list of sentences, one at a time. They were instructed to read each sentence, decide whether it was acceptable or not, and to push one of two buttons on the panel in front of them according to how they had decided. They were instructed to set up their criteria of acceptability on the basis of intelligibility rather than the basis of strict grammaticality: would they "understand the sentence if they heard it or read it?" They were requested to make the judgments as quickly and as accurately as possible.

Sentences were presented on a cathode ray tube display screen. The word READY appeared as a warning signal in the upper left corner of the screen one sec prior to the presentation of each sentence and disappeared immediately before the presentation of the sentence. The sentences were presented beginning on the middle line of the display screen and continuing on the following line if length made this necessary. Each sentence remained on the screen for a total of four seconds. During this time the subjects had to read the sentence, make their decision and push the appropriate button. The sentence disappeared at the end of the four sec interval, regardless of response. There was approximately a 1.5 sec intertrial interval. Each subject first saw 20 practice sentences of the same types as those used in the remainder of the experiment. These sentences were taken from those sentences on the pre-test which were not used as experimental sentences. After the practice sentences, each subject saw the sentences in random order. The presentation and recording of the response latencies was done by a PDP 8/I computer.

4.3.1. Subjects

Subjects were 40 University of Texas undergraduates who participated in the experiment as part of a requirement for a course in introductory psychology. They were tested two at a time or individually, according to the availability of subjects. At the conclusion of each testing session, each subject was given a written explanation of the purpose of the experiment, and any questions about the purpose or procedures used in the experiment were answered.

4.4. Results and Discussion

As in the other experiments, the performance of the subjects was not perfect with regard to their decisions. These incorrect responses were discarded, and a mean of the correct responses for each subject for each cell was computed. Table 11 shows a mean of these data for each cell. The error rates for each cell are shown in Table 12.

A quick inspection of the data in Table 11 shows that grammatical non-idioms took the least time, that idioms of both types took longer than non-idioms which were grammatical, and that ungrammatical non-idioms took the most time. The judgments accorded well with the linguistic well-formedness or the lack of it. All of the results were as predicted.

The grammaticality effect was significant, $F(2,78)=112.84$, $p<.0001$, indicating that grammatical sentences were easier to make decisions about, and that idioms were harder than grammatical non-idioms but easier than ungrammatical sequences.

The two Types showed no main effect, $F(1,39)= .089$, $p>.75$, but the interaction of this variable and the grammaticality variable was significant, $F(2,78)= 20.51$, $p<.0001$. This interaction indicates that grammaticality behaved differently within the two Types. The Type I materials show shorter response latencies than Type II in the non-idiomatic cells, but the trend is just the reverse in the idiom cells, with the Type I idioms (the ambiguous idioms) taking longer than the Type II idioms (the unambiguous ones).

Within the grammaticality variable there is another variable, that of response type. The subjects decided that the ungrammatical non-idioms were not acceptable and pushed the NO button for these kinds of sentences, whereas they decided that the grammatical non-idioms, the grammatical idioms

TABLE 11

RESULTS FROM EXPERIMENT III

	GRAMMATICAL NON-IDIOMS	IDIOMS	UNGRAMMATICAL NON-IDIOMS
TYPE I	1.906	2.265	2.399
TYPE II	1.954	2.146	2.455

TABLE 12

ERROR RATES FROM EXPERIMENT III

	GRAMMATICAL NON-IDIOMS	IDIOMS	UNGRAMMATICAL NON-IDIOMS
TYPE I	1%	5%	2%
TYPE II	2%	3%	5%

and even the ungrammatical idioms were acceptable and pushed the YES button. Some investigators (e.g., Rubenstein, Lewis & Rubenstein, 1971) have found that YES decisions are different in response latency than NO decisions, others (e.g., Posner & Mitchell, 1970) have found no such difference. In the previous experiments discussed in this paper, differences between decision types have been found. In order to be sure that any grammaticality effect found to be significant was not merely a decision type confounding, another analysis of variance was performed using only the two levels of grammaticality to which the response on acceptability was YES. The two types were retained.

Again, grammaticality (non-idiom vs. idiom) was significant, $F(1,39)=190.10$, $\underline{p}<.0001$, indicating that idioms are harder to make decisions about than non-idioms. Again, the two types did not produce a significant difference, $F(1,39)= 3.15$, $\underline{p}>.08$. And the interaction of these two variables was once more significant, $F(1,39)= 38.89$, $\underline{p}<.0001$. The presence of this interaction indicates that idiomaticity is a phenomenon separable from the grammaticality variable, yet still influenced by it.

The lack of an overall Type effect indicates that the types do not differ on the average, but the presence of an interaction indicates that there is some effect within these variables. A closer examination of the cell means clarifies this: the responses to the grammatical idioms were slower than to the ungrammatical ones, but the responses to the non-idiomatic materials show just the opposite pattern, producing a highly significant interaction but no overall main effect. There appears to be some regular difference between the two Types which appears in the non-idiomatic cells. The effect from the Types which appears in the non-idiomatic cells appears to be non-significant and is probably due to some difference in materials. This difference is obscured and even reversed in the idiom cells, strongly indicating the presence of another effect. This difference between the two idiom cells cannot be part of the same trend found in the other materials since it is in the opposite direction.

The difference between the two idiom cells is the comparison of interest, and the main motivation for this experiment. These two cells represent grammatical (ambiguous) and ungrammatical (unambiguous) idioms. Since this comparison was a planned one, a t-test was performed testing the difference between these two cells. This difference was significant, $t(39)=7.54$, $p<.001$, substantiating the claim that ungrammatical idioms are handled differently from grammatical ones in comprehension tasks.

The difference between grammatical and ungrammatical idioms found in this experiment was not in the same direction as that suggested by the results of Experiment II. The results of Experiment III seem to be the more believable because of the strict controls in this experiment and the generally unexplainable results from Experiment II. The difference found between these two conditions in this experiment is at an extro-ordinary level of significance while that in the second experiment was only a non-significant trend. The results from this experiment, then, seem to be the results which need explanation.[1]

---

[1] Another piece of evidence confirming the results from Experiment III can be found by re-examining the results from Experiment I. In Experiment I there was no division between ambiguous and unambiguous materials. However, such a post hoc division can be made. By examining all of the materials for the idiom-idiom DIFFERENT cells (in order to find materials comparable to those tested by a t-test in Experiment II), it is possible to obtain three pairs in which both sentences are ambiguous and three in which both are unambiguous. When the means for these two groups are calculated, a large difference is found. This difference is in the same direction as

To recapitulate: it takes longer to make a decision of unacceptability than one of acceptability. Within the grammaticality-type decision, it takes longer to decide that idioms are acceptable than to decide that non-idioms are acceptable. Within the idiom types, it takes longer to decide that a grammatical idiom is acceptable than to decide that an ungrammatical idiom is acceptable.

Grammaticality does seem to play a role in sentence comprehension. It does not take as long to perform a task with fully grammatical sentences as with grammatically deviant sentences. Idiomaticity does not clearly fit into either of these categories, although it has things in common with both. The subjects judged the great majority (97.5%) of the idiomatic sentences and also the non-idiomatic grammatical sentences (98%) to be acceptable. Both of these kinds of sentences are comprehensible English sentences to most people, whereas the ungrammatical non-idiomatic sentences were almost always (96.5%) judged unacceptable. However, the speed with which this task was performed revealed that idioms presented not exactly the same kind of processing problem as either grammatical or ungrammatical sentences which are not idiomatic. The time to decide on the acceptability of an idiom is longer than that of a grammatical sentence yet shorter than that of an ungrammatical sentence. Even though idioms do seem to show some effects of grammaticality, these data seem to indicate that idioms are separable from the realm of non-idiomatic sentences.

---

[1] (cont.) that found in Experiment III; that is, the ambiguous idioms took longer than the unambiguous. Unfortunately, any statistical test for significance would be inappropriate since in making a division within one cell, the control for length is destroyed. However, the difference is a very large one (600 msec) and does suggest that this result is a reliable one.

CHAPTER V

A MODEL TO EXPLAIN THE DATA

In this chapter, an attempt will be made to integrate the empirical data about idioms into a coherent model of the processes involved in understanding idioms. First, to recapitulate the results which need to be incorporated in this model: Experiment I indicated that idioms are more difficult to deal with in a similarity matching task than non-idiomatic sentences. The degree of similarity of the materials also influenced the speed with which the task was done, decisions on IDENTICAL materials being fastest, decisions on DIFFERENT materials being slower and SAME decisions on materials with semantic similarity being slowest.

The second experiment also contained the variable of similarity of materials and used the same comprehension task as the first experiment. Experiment II was different in that the variable of ambiguity was included as a separate variable whereas ambiguity was not a separate variable in Experiment I. Ambiguity was found to be a significant variable in Experiment II, with ambiguous materials requiring longer to make the similarity judgment than unambiguous materials. The ambiguity effect was most prominent with the ambiguous non-idiom materials. In the first experiment, the idioms took longer than non-idioms on the decision task; this was not the case for Experiment II, the idioms did not take as long, a reversal of the results from Experiment I. As in Experiment I, the similarity of materials was again a significant variable, with the results showing the same pattern as in the first experiment.

In an attempt to explain the reversal of the idiom effect, a bias survey was conducted. It was found that the materials used in Experiment II did have differing degrees of bias; the ambiguous lexical items (non-idioms) were less biased than the ambiguous idioms which were, in turn, less biased than the unambiguous idioms. However, the degree of bias did not correlate with the length of reaction time, indicating that, although present, the variable of bias was not totally explanatory.

The third experiment was designed to further explore the nature of the difference between the ambiguous (grammatical) and unambiguous (ungrammatical) idioms. The task used previously was unsuitable for these materials and was changed to an acceptability rating task. The materials were ungrammatical and grammatical idioms and non-idioms. The results indicated that grammatical non-idioms were accepted as good English sentences the fastest, that idioms were acceptable, but took longer to accept and that ungrammatical sentences took even longer and were judged unacceptable. The idioms took longer than acceptable non-idioms, but there was a difference in decision time between the two kinds of idioms, the grammatical (ambiguous) idioms taking longer to accept than the ungrammatical (unambiguous) idioms. The proposed model will need to explain all of these results.

The model to be presented here will elaborate the differences between the course of normal comprehension processing and the comprehension processing involved with idioms.

As previously mentioned (Sec. 1.2.2.), logic necessitates that normal comprehension must at least be attempted in the case of ungrammatical idioms. It is only when normal comprehension fails that some other additional process is initiated which results in an idiomatic interpretation. Since it is known that ungrammatical idioms are judged to be acceptable ambiguous idiom are computed, just as with any other ambiguity, but it cannot be assumed that the process of finding the two readings is the same as for other ambiguities since one reading is not a normal one. How this non-normal reading is obtained is an interesting question to which there is no answer at this point. However, it is obvious that this idiomatic reading is obtained, and some basis for deciding between the two readings must exist. Context, of course, would be one basis, and another would be bias.

The bias survey (Experiment II) indicated that the idiomatic reading was always the most likely, even for ambiguous idioms; that when a meaning is reported for such strings, the meaning is most frequently that of the idiomatic reading, even though the other non-idiomatic reading has been computed (according to the Bever, Garrett & Hurtig-type model).

Are these suggestions sufficient for the formulation of a model which will explain the obtained data? Consider the data from Experiment III. In Experiment III there were grammatical non-idioms, grammatical idioms, ungrammatical idioms and ungrammatical non-idioms. Subjects decided whether these were acceptable sentences or not. For the grammatical non-idioms, normal comprehension processing proceeds smoothly and results in a reading for the sentence. The decision that this is an acceptable sentence can be made on the basis of this one reading since the obtained reading is that of an intelligible meaningful English sentence.

The decision of acceptability cannot be made so easily for any of the other cases (this difficulty is reflected in the length of the reaction times). The assumption of serial processing for ungrammatical idioms mentioned earlier will give an adequate explanation for much of the data from Experiment III. The ungrammatical idioms must go through at least a part of the normal course of comprehension, cannot be declared acceptable on this basis, and are then processed in some other additional way which results in the idiomatic reading and, hence, in a decision of acceptability. This additional process is obviously one which can discover other, non-normal readings for a string, if such are present. Presumably this process is attempted (perhaps in addition to others) before a negative acceptability decision is made. It can be seen by examining the decision times for those strings judged unacceptable that such a search for and failure to find additional non-normal readings is quite time consuming.

This additional process is not necessarily limited to that of idiom comprehension. It is entirely reasonable to postulate that other types of non-normal comprehension processing proceed in a similar way. Such things as poetic usage, metaphors, similies, etc., could all be considered to be non-normal usage and could be handled similarly by some additional processes, though what these would be is not clear. Using a computer analogy, there could be a number of subroutines which are called when the normal program fails, and the idiom subroutine is only one of several.

This process of normal comprehension plus other non-normal processes in case of failure of the normal does seem adequate to be developed into a model, and can explain much of the data from Experiment III.

However, the length of decisions for the ambiguous idioms is still unexplained with only the serial processing model. For these data the Bever et al model of ambiguity processing is relevant and enlightening. Table 13 shows the number of possible readings which would be discovered for each type of sentence used in Experiment III. Note that the ambiguous idioms have two possible readings each, and according to the proposed

TABLE 13

NUMBER OF POSSIBLE READINGS FOR MATERIALS FROM EXP. III

	IDIOM	NON-IDIOM
GRAMMATICAL	2	1
UNGRAMMATICAL	1	0

TABLE 14

NUMBER OF POSSIBLE READINGS FOR MATERIALS FROM EXP. II

AMBIGUOUS

IDIOM-IDIOM	NON-IDIOM NON-IDIOM
4 (2 for each sentence)	4 (2 for each sentence)

UNAMBIGUOUS

IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
2	2

TABLE 15

AVERAGE NUMBER OF COMPARISONS NECESSARY FOR A MATCH

AMBIGUOUS

IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
between 1 and 2, depending on bias	2

UNAMBIGUOUS

IDIOM-IDIOM	NON-IDIOM-NON-IDIOM
1	1

TABLE 16

AVERAGE NUMBER OF COMPARISONS NECESSARY FOR A MATCH

MATERIALS FROM EXP. I

IDIOM-IDIOM	IDIOM-NON-IDIOM	NON-IDIOM-NON-IDIOM
between 1 and 2*	less than 1.5*	1

* depending on both bias and grammaticality

model, both are computed. In Experiment III there was a decision required about the acceptability of the sentences. If both readings are computed, both readings might have to be examined in order to make this decision, a process which would take some time. From the data it can be seen that this does take more time than deciding on the acceptability of the one reading found by normal comprehension processing for a grammatical non-idiom and even more time than deciding that the obtained reading for an ungrammatical idiom is acceptable, but not as much time as searching all possible readings and deciding negatively. It is not possible to decide on the basis of the data from Experiment III whether the additional time for an idiom is consumed in finding the two readings for the ambiguous idioms, or in making the decision about the two readings, or whether both processes are more lengthy than for normal sentences with only one reading.

The two factors of number of possible readings and serial processing, taken together, do seem to make a model adequate to explain all the data from Experiment III. But how will such a model do with the data from Experiment II?

Recall that the task in Experiment II was a similarity-of-meaning decision made on pairs of sentences with varying degrees of similarity. The design included both the variables of idiomaticity and ambiguity. The matching procedure necessary to make a correct decision of similarity or difference is probably both more complicated and more time consuming than the acceptability decision required in Experiment III. However, both require a complete decoding of the meaning of each sentence and some decision made on the output from those decoding operations.

It is not clear that, in order to make a decision of similarity when the two sentences are physically identical, the same kind of extensive decoding processes are required as for decisions on semantic criteria. As can be seen from the data (Table 7), IDENTICAL decisions take considerably less time than either semantically similar or semantically different materials, indicating that these decisions might be made on the basis of different information. As Posner and Mitchell (1967) and others have noted, decisions of physical identity and name identity appear to be different kinds of processes.

In order to make any decision about the semantic similarity of two sentences, at least one meaning must be obtained for each of the sentences. Then some matching procedure must occur which probably determines the similarity on the basis of semantic information. In order to make a careful decision of dissimilarity, all possible meanings must be compared. This means that, if one of the sentences is ambiguous, there are three meanings and at least two matches that must be made; and if both sentences are ambiguous, there are at least four meanings and three necessary matching operations. In order to make a careful decision of similarity, a process similar to the one described above could occur. For the sentences with semantic similarity, it might be possible to stop the matching operations as soon as a good match was found, and to report the similarity, but it is also possible that this does not happen in such a simple way. Perhaps there exists some kind of criterion for a high degree of similarity, and unless this criterion is reached, the matching operations continue until all matches are done. At that point, the decision as to similarity must be made on the basis of the imperfect matches. Suppose that a number representing 'goodness of the match' is assigned as each matching operation is done. If the match is bad, this number would be zero; but if there is some degree of similarity, it would be higher than zero. At the end of all of the possible matching operations, a decision must be made. If all the numbers are zero, a decision can be made quickly because the numbers do not have to be examined further, and dissimilarity can be reported. However, if the numbers are higher than zero, closer examination is required. If some criterion of acceptable similarity is obtained, then similarity can be reported. Semantic similarity or dissimilarity would necessitate all matching operations and assignment of 'goodness of match' numbers. Semantic similarity for any match would necessitate examination and evaluation of that match with some criterion of acceptability for similarity. This evaluation would be time consuming, but could result in a decision of similarity. This kind of model explains the differences in decision time found in both Experiments II and I for decisions involving different degrees of similarity.

Since Experiment II included the ambiguity variable, the number of possible readings for each sentence should be considered. -Table 14 shows this. Due to the nature of the matching procedure, number of possible readings is directly reflected in the number of matching operations which are necessary before an accurate decision can be made. This is shown in Table 15. This number of matching operations will be influenced by bias since bias may influence how many readings are used in the matching operations for each item.

All ambiguities are influenced by bias to some degree. Bias varies not only across ambiguities, but also across people. This degree of bias indicates, in practice, just how ambiguous any item actually is, whether it has, for any individual, one or more than one meaning. For most of the idioms the bias scores indicated that the items were actually not very ambiguous; that is, the non-idiomatic reading was given so infrequently as to indicate that, in practice, the item was not ambiguous, even though in theory it was ambiguous. For some idioms this was not true--they were effectively ambiguous. But as a group, the ambiguous idioms were not very ambiguous; most people gave the idiomatic reading, suggesting that these meanings are the strongest ones. The degree of bias for the non-idiomatic ambiguities was much different in that many more items in that group were effectively ambiguous; both meanings were given. This should make for a different number of matching operations necessary (on the average) before a correct decision can be made on the ambiguous non-idioms. If only one reading (the idiom reading) is considered for each of the idiomatic items, then fewer comparisons are necessary than if all readings are compared with each other. This could happen even though all readings are obtained initially. Cairns (1973) and also Swinney (1974) have suggested that all readings of an ambiguity are initially computed, but that the length of time that elapses before one of these is selected is a function of the relative strength of the readings. Thus, a decision is made more quickly for a biased ambiguity than for an unbiased one. It is possible that both the idiomatic and non-idiomatic reading for an ambiguous idiom are retrieved, but only the idiomatic reading would be retained at the time of the matching operation. As suggested by Cairns and by Swinney, in the case where the strengths of the two readings do not differ strongly (the unbiased cases), both readings would be retained until some basis for determining the correct reading could be obtained.

On this account, the idioms would be functionally less ambiguous than the non-idioms; subjects would be likely to retain only the idiomatic reading for idioms whereas they would be more likely to retain both readings or non-idiomatic (unbiased) ambiguities. The difference in average number of matching operations done on the retained readings would thus be reflected in the decision-time difference between the ambiguous idioms and the ambiguous non-idioms; the idioms being faster in decision time than the less biased non-idioms. This account explains the difference between the ambiguous non-idioms and the unambiguous non-idioms; ambiguities require more comparisons since, for most people, neither meaning is strongly biased, both readings are retained, and all the comparisons are made, taking longer than the one comparison for the unambiguous cases.

The length of time required for the unambiguous idioms can again be explained by the serial processing required to get any kind of reading for these strings. There is possibly even a greater degree of difficulty in making a decision of similarity of meaning on two strings neither of which is a well-formed sentence.

The model does hold up well in explaining the results from Experiment II. Then, surely, Experiment I will fit in with this model.

Table 16 shows the average number of matching operations necessary for a correct decision. Since there was no separation of materials according to ambiguity, the idiom-idiom cells contained idioms which were ambiguous and idioms which were ungrammatical. Therefore, the serial processing effect as well as the number of comparisons necessary for a decision are both working in these cells and are both affecting decision time, which is large. These same effects are working in the idiom-non-idiom cells, but for only the idiom half of each pair. This means that two is the maximum number of matching operations which can be made and one is the minimum number. But considering the bias factor, an average of 1.5 comparisons is probably an overestimate. For the non-idiom pairs, only one comparison is possible, and at least that is necessary.

Since the bias of the items used in Experiment I is not known, an accurate estimate of the number of comparisons necessary is not available. However, it seems safe to say on the basis of the bias survey for Experiment II that there is some bias affecting the comparisons involving idioms.

The proposed model does account for much of the data on idioms. It accounts for the data from Experiment I: the length of decisions involving idioms. It accounts for the data from Experiment II: the difference between ambiguous and unambiguous non-idioms, and the reversal of the idiom effect found in Experiment I. The proposed model would explain the differences in decision time found for the three similarity conditions in both the first and second experiments. The data from Experiment III are also explained: the length of decisions for grammatical and ungrammatical idioms and non-idioms.

There are many questions about idioms to which the model does not speak. Some of these questions are unanswered because the data which would answer them were not part of the experiments and, therefore, remain uncollected. The other questions are unanswered because the model is not fully explanatory.

One such question is the question of triggering of an idiomatic reading. With an ambiguous grammatical string, one reading of which is an idiom, the question of triggering the idiomatic reading is certainly an interesting one. As previously mentioned, it is not necessary for present purposes to answer this question. We need only assume that somehow the idiomatic reading of an ambiguous idiom is obtained. But we would have a more complete model of idiom comprehension if we could. The question is vital to a full understanding of the comprehension processes involved in understanding idioms. The proposed model does not explain how the idiomatic reading is found; it only hypothesizes that, as in the case of other ambiguities, both (or all) readings of the ambiguity are found. As previously mentioned there is reason to believe that this process may be different for idiomatic ambiguities than for non-idiomatic ambiguities.

The question of what initiates the search for the idiomatic reading is an intriguing one. It could be accomplished by the appearance of some key word of the idiom string, such as 'bucket' in 'kick the bucket'. This would have to be a context-sensitive arrangement where 'bucket' would have to appear in the presence of the other part of the idiom, that is only following the words 'kick the'. If this formulation is of the correct sort, this would imply some features idioms must have. It would require that the precise string occur with no deviations. This is accurate for some idioms; whereas the a) part of Sentence (12) is idiomatic, b) and c) are not. However, for some idioms, this is not desirable, since a degree of lexical substitution is allowed with retention of idiomaticity, as in Sentence (13). A context-sensitive formulation could not allow for lexical substitution unless the context was sensitive to semantic information (selectional restrictions); and the problem of how such information might be used is a sticky one. This seems to leave some degree of idiosyncrasy

(12). a) John kicked the bucket.
b) John kicked the pail.
c) John booted the bucket.

(13). a) Bill was ready to throw in the towel.
b) Bill was ready to toss in the towel.

in the behavior of idioms which would require each idiom as well as its permitted variations to be stored along with the triggering context for each. It does not permit the formulation of a general rule, but rather a list upon which the machinery of comprehension must operate. The proposed model does not determine the nature of the process, but it does entail some process which accomplishes an idiomatic reading for such ambiguities.

The second problem with the model is the separation of the processes necessary for the comprehension of grammatical and ungrammatical idioms. According to the proposed model, these two processes are handled by completely separate mechanisms. This is infelicitous for two reasons. First, the complete separation of the mechanisms makes it impossible to make predictions about the nature of the relationship between the two kinds of idioms, that is, it is not possible to predict which will take the longest amount of time in any task, the computation of all possible comparisons or the serial computation of the various non-normal subroutines to find some reading. The relationship of the two types of idioms with respect to each other in comprehension tasks is not predictable. The second is the theoretical extension of the first: if such processes are hypothesized as completely separate, there seems to be absolutely no relationship between the two types of idioms, and, since no relationship exists between them, calling both idioms is not an accurate classification. As noted in Sec. 1.4., Chafe has criticized the linguistic descriptions of idioms for precisely this reason; they made separate provision for the different kinds of idioms. Intuitively, this does not seem to be true. Idioms, whether ambiguous or ungrammatical, do seem to be part of one phenomenon. Both kinds of idioms do share the lack of normality of a reading; for the ambiguous idioms, one of the readings is of a non-compositional nature, and for the ungrammatical idioms, the only possible reading is. It is possible to incorporate this into the model and thereby establish a relationship between the two kinds of idioms. The separate type of mechanisms which trigger the idiomatic readings for grammatical and ungrammatical idioms need not also necessitate that the kinds of information available about the idiomatic readings also be different. Perhaps, once triggered, the processes which would 'look up' an idiomatic reading are the same for both kinds of idioms. The data do support some kinds of difference between these two kinds of sentences which people call idioms, but the difference could lie entirely in the kinds of triggering mechanisms for each, and does not necessitate complete independence.

The proposed model does seem to fit well with the available data, as well it should. There are other data which are not yet available which would serve to test the model. One type of data could be collected in a paradigm very similar to that used for Experiment III. Recall that in Experiment III there were grammatical and ungrammatical idioms and non-idioms and all of the grammatical non-idioms were unambiguous. It took less time to decide that these grammatical non-idioms were acceptable sentences than to decide that either kinds of idiom was acceptable and to decide that the ungrammatical non-idioms were unacceptable. In this kind of paradigm, the model explains the length of decision time for the ungrammatical strings as being caused by the time taken to complete a serially ordered series of processes which finds a non-normal reading if there is one, or determines that there is no possible reading. The decision time for the ambiguous idioms is explained by the time necessary to find both readings and to decide on the acceptability of both. Bias must play a role in this process, as with any kind of ambiguity. If this model is correct, other ambiguities would also cause decision difficulty, especially if the ambiguities were unbiased ambiguities. If this model is correct, these unbiased non-idiomatic ambiguities should cause more difficulty in the decision task than the ambiguous idioms, due to the difference in bias. The model would predict the same kind of idiom-non-idiom difference found in Experiment II: the lexical ambiguities should take longer than the idiom ambiguities.

In addition, this model would predict another effect of ambiguity; a difference between unbiased and biased non-idiomatic ambiguities. The biased ambiguities should be easier than the unbiased ambiguities due to the difference in the average number of matching operations required.

This would seem to constitute two valid (and easily accomplished) tests of the proposed model. If the predictions are confirmed by the data, then the model would seem worthy of further elaboration. This model is not only a possible model of idiom comprehension, but an elaboration of a model for ambiguity comprehension as well as a basis for a model of all non-normal comprehension processes. Some basis for further thought and research on non-normal comprehension would certainly be a valuable addition to the area of comprehension research.

As previously mentioned, there are questions about idiom comprehension which remain unanswered by the model and by the data. The question of the process which initiates the search for an ambiguous idiom is an intriguing one, as is the question of the processes involved in finding the non-compositional readings for either sort of idiom. These questions, along with those about non-normal comprehension in general deserve further investigation.

The proposed model should be considered only a very tentative one which may merit further elaboration of the basis of further data. The model is, admittedly, horrendously complex and is based on incomplete data. But the available data are also very complex. Simplicity is esthetically pleasing, but in this case it does not seem descriptively adequate.

APPENDIX A

MATERIALS FROM EXPERIMENT I

SEMANTIC SIMILARITY, IDIOM-IDIOM

Get on the ball. Get on the stick.

He popped his top. He blew a fuse.

Buzz off. Bug off.

Go jump in the lake. Go fly a kite.

They were pulling John's leg. They were putting John on.

He went nuts. He went bananas.

He really got a kick out of that. He really got a bang out of that.

He lost his mind. He lost his marbles.

He blew his top. He flew off the handle.

He let the cat out of the bag. He spilled the beans.

Get lost. Beat it.

The teacher put her foot down. The teacher laid down the law.

SEMANTIC SIMILARITY, IDIOM-NON-IDIOM

Get on the ball. Be more alert.

He blew a fuse. He became angry.

Buzz off. Go away.

Go jump in the lake. Get out of here.

They were putting John on. They were teasing John.

He went nuts. He became insane.

They really got a bang out of that. They really enjoyed that a lot.

He became insane. He lost his mind.

He blew his top. He became angry.

He let the cat out of the bag. He told about the surprise.

Get lost. Go away.

The teacher laid down the law. The teacher enforced the rules.

SEMANTIC SIMILARITY, NON-IDIOM-NON-IDIOM

Be more alert. Start thinking.

He became enraged. He became angry.

Go away. Leave.

Get out of here. Get away from me.

They were teasing John. They were kidding John.

He became deranged. He became insane.

They really enjoyed that a lot. They really liked that very much.

He lost mental control. He became insane.

He became angry. He reacted angrily.

He told about the surprise. He divulged the secret.

Go away. Leave.

The teacher became more strict. The teacher enforced the rules.

DIFFERENT, IDIOM-IDIOM

Go jump in the lake. Get on the ball.

He blew a fuse. He made the scene.

Buzz off. Be cool.

Go jump in the lake. Break that habit.

They were putting John on. He turned over a new leaf.

He went nuts. He drew a blank.

The old man kicked the bucket. They really got a bang out of that.

He really got my goat. He lost his mind.

He blew his top. He stood a chance.

That's only a drop in the bucket. He let the cat out of the bag.

Think quick. Get lost.

The teacher laid down the law. The teacher flew the coop.

DIFFERENT, IDIOM-NON-IDIOM

Get on the ball. He needs new shoes.

He blew a fuse. The rain fell.

Buzz off. He left.

Go jump in the lake. Jan did mail the letter.

They were putting John on. We decided to do our work.

He went nuts. She was on time.

They really got a bang out of that. Alice looked at the magazine.

He lost his mind. The plan did work.

He blew his top. Gary found him.

He let the cat out of the bag. The cement floor was cold.

Get lost. She danced.

The teacher laid down the law. Leopards are becoming extinct.

DIFFERENT, NON-IDIOM-NON-IDIOM

Start thinking. A friend came by.

He became enraged. He moved the table.

Leave. Come in.

Get away from me. The dog hurt its paw.

Ellen found only two cups. They were kidding John.

He became deranged. The card was blank.

They really liked that very much. The young man bought a coat.

He smashed the vase. He lost mental control.

He reacted angrily. She lost her ring.

He divulged the secret. His gloves are not here.

Leave. He died.

The teacher enforced the rules. The phone number was unlisted.

IDENTICAL, IDIOM-IDIOM

Get on the stick. Get on the stick.

He popped his top. He popped his top.

Bug off. Bug off.

Go fly a kite. Go fly a kite.

They were pulling John's leg. They were pulling John's leg.

He went bananas. He went bananas.

They really got a kick out of that. They really got a kick out of that.

He lost his marbles. He lost his marbles.

He flew off the handle. He flew off the handle.

He spilled the beans. He spilled the beans.

Beat it. Beat it.

IDENTICAL, NON-IDIOM-NON-IDIOM

Start thinking. Start thinking.

He became enraged. He became enraged.

Leave. Leave.

Get away from me. Get away from me.

They were kidding John. They were kidding John.

He became deranged. He became deranged.

They really liked that very much. They really liked that very much.

He lost mental control. He lost mental control.

He reacted angrily. He reacted angrily.

He told about the surprise. He told about the surprise.

Leave. Leave.

The teacher became more strict. The teacher became more strict.

APPENDIX B

MATERIALS FROM EXPERIMENT II

AMBIGUOUS MATERIALS

*SEMANTICALLY SIMILAR, IDIOM-IDIOM

That made me shake in my shoes. That made me tremble like a leaf.

He spilled the beans. He let the cat out of the bag.

The old man kicked the bucket. The old man turned up his toes.

Bill paved the way for him. Bill cleared the road for him.

They were pulling John's leg. They were handing John a line.

We are in the same boat. We are in the same fix.

They said he had cold feet. They said he was yellow.

He was skating on thin ice. He was playing with fire.

He lost his marbles. He has a screw loose.

He was afraid to break the ice. He was afraid to take the first step.

He told Bill to go fly a kite. He told Bill to go jump in the lake.

**SEMANTICALLY SIMILAR, IDIOM-NON-IDIOM

That made me shake in my shoes. That frightened me very much.

He spilled the beans. He told the secret.

The old man kicked the bucket. The old man died.

Bill paved the way for him. Bill made things easy for him.

They were pulling John's leg. They were teasing John.

We are in the same boat. We are in the same situation.

They said he had cold feet. They said he was a coward.

He was skating on thin ice. He was in a bad situation.

He lost his marbles. He became insane.

He was afraid to break the ice. He was afraid to begin.

He told Bill to go fly a kite. He told Bill to get out of here.

**SEMANTICALLY SIMILAR, NON-IDIOM-NON-IDIOM

That frightened me very much. That made me very afraid.

He told the secret. He revealed the story.

The old man died. The old man perished.

Bill made things easy for him. Bill prepared things for him.

They were teasing John. They were kidding John.

We are in the same situation. We are in the same predicament.

They said he was a coward. They said he was afraid.

He was in a bad situation. He was in a lot of trouble.

He became insane. He became crazy.

He was afraid to begin. He was afraid to start.

He told Bill to get out of here. He told Bill to go away.

DIFFERENT, IDIOM-IDIOM

That made me shake in my shoes. That slipped through my fingers.

He spilled the beans. He wouldn't play ball.

The old man kicked the bucket. That's just a drop in the bucket.

Bill paved the way for him. Bill often hit the bottle.

They were pulling John's leg. They turned over a new leaf.

We are in the same boat. We could pave the way for him.

Thay said he had cold feet. They made both ends meet.

He was skating on thin ice. He kept me in the dark.

He lost his marbles. He spilled the beans.

He was afraid to break the ice. He was really out on a limb.

He told Bill to go jump in the lake. He bit off more than he could chew.

*DIFFERENT, IDIOM-NON-IDIOM

That made me shake in my shoes. There is no mint in this city.

He spilled the beans. He inspected the tanks.

The old man kicked the bucket. The bats were found in the attic.

Bill paved the was for him. Bill inspected the tanks.

They were pulling John's leg. The farmer had a bad fall.

We are in the same boat. The mold is still wet.

They said he had cold feet. The guide was inaccurate.

He was skating on thin ice. She cut her hand on the foil.

He lost his marbles. He examined the case.

He was afraid to break the ice. He cut his hand on the foil.

He told Bill to go jump in the lake. The shop sent the iron yesterday.

DIFFERENT, NON-IDIOM-NON-IDIOM

There is no mint in this city. There is no one to make the toast.

He inspected the tanks. He got a good tip.

The cold made John miserable. The bats were found in the attic.

Bill inspected the tanks. Bill's speech was not clear.

The farmer had a bad fall. They looked through the records.

The mold is still wet. The meal was quite tasty.

The guide was inaccurate. The boxer ran around the block.

She couldn't bear children. She cut her hand on the foil.

He examined the case. He cleaned the pipes.

He cut his hand on the foil. He had a flat in downtown London.

The shop sent the iron yesterday. The roar of the fans deafened us.

IDENTICAL, IDIOM-IDIOM

That made me shake in my shoes. That made me shake in my shoes.

He spilled the beans. He spilled the beans.

The old man kicked the bucket. The old man kicked the bucket.

Bill paved the way for him. Bill paved the way for him.

They were pulling John's leg. They were pulling John's leg.

We are in the same boat. We are in the same boat.

They said he had cold feet. They said he had cold feet.

He was skating on thin ice. He was skating on thin ice.

He lost his marbles. He lost his marbles.

He was afraid to break the ice. He was afraid to break the ice.

He told Bill to go jump in the lake. He told Bill to go jump in the lake.

IDENTICAL, NON-IDIOM-NON-IDIOM

There is [illegible] one to make the toast. There is no one to make the toast.

He got a very good tip. He got a very good tip.

The cold made John miserable. The cold made John miserable.

Bill's speech was not clear. Bill's speech was not clear.

They looked through the records. They looked through the records.

The meal was quite tasty. The meal was quite tasty.

The boxer ran around the block. The boxer ran around the block.

She couldn't bear children. She couldn't bear children.

He cleaned the pipes. He cleaned the pipes.

He had a flat in downtown London. He had a flat in downtown London.

The roar of the fans deafened us. The roar of the fans deafened us.

UNAMBIGUOUS MATERIALS

*SEMANTICALLY SIMILAR, IDIOM-IDIOM

The old house makes my flesh creep. The old house curdles my blood.

She gave him the brush-off. She gave him the cold shoulder.

He skipped town before the trial. He blew town before the trial.

George struck it rich. George hit the jackpot.

They were putting John on. They were tooling John around.

He snapped Fred's head off. He jumped down Fred's throat.

That makes my blood boil. That makes me see red.

His story doesn't hold water. His story doesn't ring true.

He went bananas. He went off his nut.

My refrigerator is out of whack. My refrigerator is on the blink.

She went to pieces over the news. She fell apart over the news.

*SEMANTICALLY SIMILAR, IDIOM-NON-IDIOM

The old house curdles my blood. The old house frightens me.

She gave him the brush-off. She was discourteous to him.

He skipped town before the trial. He left town before the trial.

George struck it rich. George won some money.

They were putting John on. They were teasing John.

He jumped down Fred's throat. He spoke angrily to Fred.

That makes my blood boil. That makes me very angry.

His sotry doesn't hold water. His story doesn't seem true.

He went bananas. He became insane.

My refrigerator is out of whack. My refrigerator is not working.

She went to pieces over the news. She lost control when she heard the news.

*SEMANTICALLY SIMILAR, NON-IDIOM-NON-IDIOM

The old house frightens me. The old house makes me afraid.

She was discourteous to him. She treated him quite rudely.

He left town before the trial. He departed before the trial.

George won some money. George won a fortune.

They were teasing John. They were kidding John.

He spoke angrily to Fred. He yelled at Fred while angry.

That makes me very angry. That thoroughly enrages me.

His story doesn't seem true. His story doesn't seem accurate.

He became insane. He became crazy.

My refrigerator is not working. My refrigerator is not running.

She lost control when she heard the news. She became upset when she heard the news.

DIFFERENT, IDIOM-IDIOM

The old house curdles my blood. The small boy told a tall story.

She gave him the brush-off. She had to eat her words.

He skipped town before the trial. His teacher laid down the law.

George struck it rich. George gets under my skin.

They were putting John on. They were ready to lend an ear.

He jumped down Fred's throat. He bowed to Jim's wishes.

That makes my blood boil. That stretches the imagination.

His story doesn't hold water. He got a bang out of that.

He went bananas. He laid an egg.

My refrigerator is out of whack. My aunt hauled him over the coals.

She went to pieces over the news. She eventually had to eat her words.

*DIFFERENT, IDIOM-NON-IDIOM

The old house curdles my blood. The teacher put her hat down.

She gave him the brush-off. She decided to keep the cat.

He skipped town before the trial. She bought a pair of rollerskates.

George struck it rich. George's friend came by.

They were putting John on. They decided to do the work.

He jumped down Fred's throat. He wanted to smoke his cigar.

That makes my blood boil. His gloves are in there.

His story doesn't hold water. The young man bought a coat.

He went bananas. He hid the money.

My refrigerator is out of whack. My father criticized the election.

She went to pieces over the news. She saw the horse after he did.

DIFFERENT, NON-IDIOM-NON-IDIOM

The teacher put her hat down. The student turned in her paper.

She decided to keep the cat. She got a severe sore throat.

She was trying to put her coat on. She bought a pair of rollerskates.

George's friend came by. George's gloves are here.

He looked through the window. He found only two of the cups.

He wanted to smoke his cigar. He complained about the food.

His gloves are not here. His curtains are torn.

He bought a new suede coat. He lost his blue sweater.

He hid the money. His tire is flat.

My father criticized the election. My sister likes her history professor.

She saw the horse after he did. She looked closely at the photo.

IDENTICAL, IDIOM-IDIOM

The old house curdles my blood. The old house curdles my blood.

She gave him the brush-off. She gave him the brush-off.

He skipped town before the trial. He skipped town before the trial.

George struck it rich. George struck it rich.

They were putting John on. They were putting John on.

He jumped down Fred's throat. He jumped down fred's throat.

That makes my blood boil. That makes my blood boil.

His story doesn't hold water. His story doesn't hold water.

He went bananas. He went bananas.

My refrigerator is out of whack. My refrigerator is out of whack.

She went to pieces over the news. She went to pieces over the news.

IDENTICAL, NON-IDIOM-NON-IDIOM

The student turned in her paper. The student turned in her paper.

She got a severe sore throat. She got a severe sore throat.

She was trying to put her coat on. She was trying to put her coat on.

George's gloves are here. George's gloves are here.

He looked through the window. He looked through the window.

He complained about the food. He complained about the food.

His curtains are torn. His curtains are torn.

He lost his blue sweater. He lost his blue sweater.

He hid the money. He hid the money.

My sister likes her history professor. My sister likes her history professor.

She looked closely at the photo. She looked closely at the photo.

*Included in the experiment, but not in the analysis
**Not conforming to the criteria of ambiguity

APPENDIX C

MATERIALS AND BIAS SCORES FROM THE BIAS SURVEY

Bias score indicates the proportion of people giving each meaning.

SENTENCE TYPE	BIAS SCORE
LEXICAL AMBIGUITIES	
He cleaned the pipes.	
a. a long tube used to conduct liquid, gas, etc.	.36
b. a device used for smoking	.46
c. other meanings	.18
He got a very good tip.	
a. a fee or gratuity given for service	.50
b. a hint	.50
They looked through the records.	
a. phonograph records	.49
b. written documents	.51
Bill's speech was not clear.	
a. enunciation	.49
b. in content	.51
The mold is still wet.	
a. a fungus growth	.20
b. a cavity in which a substance is shaped	.51
c. other	.20
The cold made John miserable.	
a. minor illness	.42
b. temperature	.58
He had a flat in downtown London.	
a. tire which lost air	.40
b. small apartment	.60
The guide was inaccurate.	
a. person who conducted the tour	.32
b. instruction book	.68
The boxer ran around the block.	
a. fighter	.32
b. dog	.68

There is no mint in this city.

a.	aromatic herb	.28
b.	place where coins are manufactured	.72

He examined the case.

a.	facts	.26
b.	container	.74

Bill inspected the tanks.

a.	armoured vehicles	.24
b.	storage containers	.76

She couldn't bear children.

a.	was sterile	.22
b.	did not like	.78

He cut his hand on the foil.

a.	piece of fencing equipment	.20
b.	thin aluminum	.80

The roar of the fans deafened us.

a.	machines for producing air current	.20
b.	enthusiasts	.80

The bats were found in the attic,

a.	small nocturnal mammals	.19
b.	pieces of wood with which to hit a ball	.81

The shop sent the iron yesterday.

a.	implement used to press clothing	.15
b.	metal	.85

The farmer had a bad fall.

a.	autumn	.06
b.	accident	.94

The meal was tasty.

a.	grain	.00
b.	things to eat	1.00

AMBIGUOUS IDIOMS

He was skating on thin ice.

a.	skating on ice that was not thick	.50
b.	in a dangerous situation	.50

That slipped through my fingers.

a.	dropped it	.44
b.	got away	.46
c.	other	.10

He wouldn't play ball.	
a. play a game with a ball	.50
b. co-operate	.50
He was afraid to break the ice.	
a. crack the frozen water	.18
b. start	.82
They saud he had cold feet.	
a. feet which were cold	.18
b. was a coward	.82
They barely made both ends meet.	
a. brought the ends together	.16
b. had enough money	.84
He was really out on a limb.	
a. in a tree	.10
b. in trouble	.90
He spilled the beans.	
a. spilled some beans from a container	.10
b. revealed a secret	.90
He was afraid to take the first step.	
a. to move his feet	.10
b. start	.90
He kept me in the dark.	
a. in an unlighted place	.08
b. uninformed	.92
The old man kicked the bucket.	
a. kicked a pail	.08
b. died	.92
He lost his marbles.	
a. lost his toys	.06
b. went crazy	.94
Bill paved the way for him.	
a. asphalted a road	.06
b. made things easy	.94
He bit off more than he could chew.	
a. took a bite which was too large	.06
b. attempted a task which he was incapable of accomplishing	.94

Bill often hit the bottle.	
a. struck a bottle	.06
b. drank	.94
That's just a drop in the bucket.	
a. a drop of liquid in a bucket	.04
b. a minute amount	.96
They were pulling John's leg.	
a. tugging on a lower limb	.02
b. teasing	.98
We are in the same boat.	
a. occupy the same boat	.02
b. in the same predicament	.98
They turned over a new leaf.	
a. moved a leaf	.02
b. started anew	.98
He told Bill to go jump in the lake.	
a. get in the lake	.02
b. go away	.98
That made me shake in my shoes.	
a. move	.00
b. afraid	1.00

UNAMBIGUOUS IDIOMS

He laid an egg.	
a. produced an egg.	.40
b. talked unsuccessfully	.60
My aunt hauled him over the coals.	
a. tortured with heat	.30
b. scolded severely	.70
His story doesn't hold water.	
a. is porous	.06
b. is not true	.94
George struck it rich.	
a. hit something	.06
b. became wealthy	.94
That stretches the imagination.	
a. pulls the imagination out of physical shape	.04
b. is hard to believe	.96

The old house curdles my blood.	
a. causes blood to curdle	.04
b. frightens	.96
He went bananas.	
a. went to a place with bananas	.04
b. went crazy	.96
The teacher laid down the law.	
a. put a law down	.04
b. became more strict	.96
He jumped down Fred's throat.	
a. jumped into Fred's mouth	.04
b. became angry with Fred	.96
She gave him the brush-off.	
a. gave him an object	.02
b. was rude to him	.98
That makes my blood boil.	
a. raises the temperature of my blood	.02
b. makes me very angry	.98
The small boy told a tall story.	
a. a large story	.02
b. an exaggerated story	.98
They were ready to lend an ear.	
a. give the temporary use of an ear	.02
b. listen	.98
She went to pieces over the news.	
a. came apart into fragments	.02
b. became very upset	.98
He bowed to Jim's wishes.	
a. make a bowing movement	.02
b. complied with	.98
My refrigerator is out of whack.	
a. out of a place	.00
b. not working properly	1.00
She had to eat her words.	
a. consume words	.00
b. retract a previous statement	1.00

APPENDIX D

MATERIALS FROM EXPERIMENT III

TYPE I

GRAMMATICAL (AMBIGUOUS) IDIOMS

This could continue until the cows come home.

He was ready to give up the ship.

He often hit the bottle.

We aren't out of the woods yet.

He took the bull by the horns.

He spilled the beans.

He was out on a limb.

He let the cat out of the bag.

The old man kicked the bucket.

He was down in the dumps.

That is not my cup of tea.

He missed the boat.

We buried the hatchet.

We caught him napping.

He was afraid of his own shadow.

I knew that I should hold my tongue.

GRAMMATICAL NON-IDIOMS

This should stop before the child becomes ill.

He was ready to look up the date.

He nearly broke the clock.

We aren't on top of the hill yet.

He saw the package by the lamp.

He smashed the jars.

He was out at a museum.

He took the cup out of the box.

The short man repaired the tray.

He was down by the pier.

This isn't his piece of pie.

He rode the train.

He drew the diagram.

She took him visiting.

He was already with his first guest.

I wished that I could water my flowers.

UNGRAMMATICAL NON-IDIOMS

This should possess until the case goes over.

He was ready to feed in the nephew.

He often hit the breath.

We aren't off in the feet yet.

He looked the stem by the nest.

He spilled the priest.

He was out on a belief.

He let the dog out of the hope.

The old man kicked the esteem.

He was down in the pang.

This is not my copy of milk.

He missed the none.

We buried the frenzy.

She caught him obtaining.

He was afraid of his second pie.

I knew that I should talk my circle.

TYPE II

UNGRAMMATICAL (UNAMBIGUOUS) IDIOMS

You had better get your ass in gear.

We thought she would break her word.

She gave him the brush-off.

He skipped town.

They were putting Bill on.

He didn't stand a chance.

John hit the jackpot.

George struck it rich.

He threw a fit.

The gangster ratted on his partner.

She fell apart over the news.

The band took a breather.

My refrigerator is out of whack.

He went bananas.

He tried to stir up trouble.

Jerry was under the weather.

GRAMMATICAL NON-IDIOMS

You had better get your camera in focus.

We wished she could sell her house.

She took him the orange.

He scrambled eggs.

He was taking Suzanne out.

He didn't save a dollar.

Mary boiled the noodles.

Bill advised him often.

He sold a box.

The bridesmaid sobbed into her pillow.

She went blind over the years.

The clerk repaired a bike.

My grocer is out of mayonnaise.

He played boldly.

He wanted to cast out spirits.

Jim was behind the church.

UNGRAMMATICAL NON-IDIOMS

You had better put your pail in hairy.

We found she might wonder her stick.

She sold him the someone.

He washed town.

They were thinking Joe on.

He hadn't studied a fear.

John fed the asterisk.

Sam aimed him nearly.

He turned a law.

The goblet peered at his commander.

She went calm over the skin.

The guest had a senility.

My refrigerator is out of dune.

He left terrific.

He tried to mix in centuries.

Mark was under the industry.

BIBLIOGRAPHY

Bever, T., Garrett, M. & Hurtig, R. The interaction of perceptual processes and ambiguous sentences. Memory & Cognition, 1973, 1, 277-286.

Bobrow, S., & Bell, S. On catching on to idiomatic expressions. Memory & Cognition, 1973, 1, 343-346.

Cairns, H. Ambiguous sentence processing. Unpublished doctoral dissertation, University of Texas at Austin, 1970.

Cairns, H. The effects of bias on the processing and reprocessing of lexically ambiguous sentences. Journal of Experimental Psychology, 1973, 97, 337-343.

Cairns, H. Confirmation of the "Early decision hypothesis" and a comparison of psycholinguistic tasks. In preparation.

Carey, P., Mehler, J. & Bever, T. When do we compute all the meanings of an ambiguous sentence? In G.B. Flores D'Arcais & W.Jm. Levelt (Eds.), Advances in Psycholinguistics. New York: American Elsevier, 1970, Pp. 61-75.

Chafe, W. Idiomaticity as an anomaly in the Chomskyan paradigm. Foundations of Language, 1968, 4, 109-125.

Chomsky, N. Syntactic Structures. Mouton: The Hague, 1957.

Chomsky, N. Degrees of grammaticalness. In J. Fodor & J. Katz (Eds.), Readings in the Philosophy of Language. Englewood Cliffs, New Jersey: Prentice Hall, 1964, Pp. 384-389.

Chomsky, N. Aspects of the Theory of Syntax. Cambridge, Mass.: M.I.T. Press, 1965.

Collins, V. A Book of English Idioms. London: Logmans, Green & Company, 1958.

Collins, V. A Second Book of English Idioms. London: Logmans, Green & Company, 1958.

Danks, J. Some factors involved in the comprehension of deviant English sentences. Unpublished doctoral dissertation, Princeton University, 1968.

Downey, R. & Hakes, D. Some psychological effects of violating linguistic rules. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 158-161.

Foss, D. Some effects of ambiguity upon sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 699-706.

Foss, D., Bever, T. & Silver, M. The comprehension and verification of ambiguous sentences. Perception and Psychophysics, 1968, 4, 304-306.

Foss, D & Jenkins, C. Some effects of context on the comprehension of ambiguous sentences. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 577-589.

Fraser, B. Idioms within a transformational grammar. Foundations of Language, 1970, 6, 22-42.

Garrett, M. Does ambiguity complicate the perception of sentences? In G.B. Flores D'Arcais & W.Jm. Levelt (Eds.), Advances in Psycholinguistics. New York: American Elsevier, 1970, Pp. 48-60.

Katz, J. Semi-sentences. In J. Katz & J. Fodor (Eds.), Readings in the Philosophy of Language. Englewood Cliffs, New Jersey: Prentice Hall, 1964, Pp. 400-416.

Katz, J. & Postal, P. Semantic interpretation of idioms and sentences containing them. Quarterly Progress Report, M.I.T. Research Laboratory of Electronics, 1963, 70, 275-282.

Lakoff, G. Pre-supposition and relative well-formedness. In D. Steinberg & L. Jakobovits (Eds.), Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology. New York: Cambridge University Press, 1971, Pp. 329-340.

MacKay, D. To end ambiguous sentences. Perception and Psychophysics, 1966, 1, 426-436.

MacKay, D. & Bever, T. In search of ambiguity. Perception and Psychophysics, 1967, 2, 193-200.

Mehler, J. & Carey, P. The interaction of veracity and syntax in the processing of sentences. Perception and Psychophysics, 1968, 3, 109-111.

Newmeyer, F. The insertion of idioms. In P. Peranteau, J. Levi & G. Phares (Eds.), Papers from the Eighth Regional Meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society, 1972, Pp. 294-302.

Posner, M. & Mitchell, R. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Rubenstein, H, Lewis, S. & Rubenstein, M. Homographic entries in the internal lexicon: effects of systematicity and relative frequency of meanings. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 57-62.

Swinney, D. Effects of context upon the processing of lexical ambiguities during sentence comprehension. Unpublished doctoral dissertation, University of Texas at Austin, 1974.

Thorndike, E. & Lorge, T. The Teacher's Wordbook of 30,000 Words. New York: Teacher's College, Columbia University, 1944.

Weinreich, U. Problems in the analysis of idioms. In J. Puhvel (Ed.), Proceedings of the Summer 1966 Linguistic Forum at the University of California, Los Angeles. Berkely, California: University of California Press, 1967, Pp. 23-81.

VITA

Linda Loraine Brannon was born in Dallas, Texas, on June 1, 1948, the daughter of Edna Vaughan Brannon and Roy Clifford Brannon. She graduated as valedictorian of the class of 1966 at Duncanville High School in Duncanville, Texas. In the fall of 1966 she entered The University of Texas at Austin. She received the degree of Bachelor of Arts with high honors with a major in psychology from The University of Texas at Austin in May of 1970. In the fall of the same year she entered the Graduate School of The University of Texas in pursuit of the degree of Doctor of Philosophy in the field of psychology.

This dissertation was typed by Carole Stratton.